

Fig. 1a-f

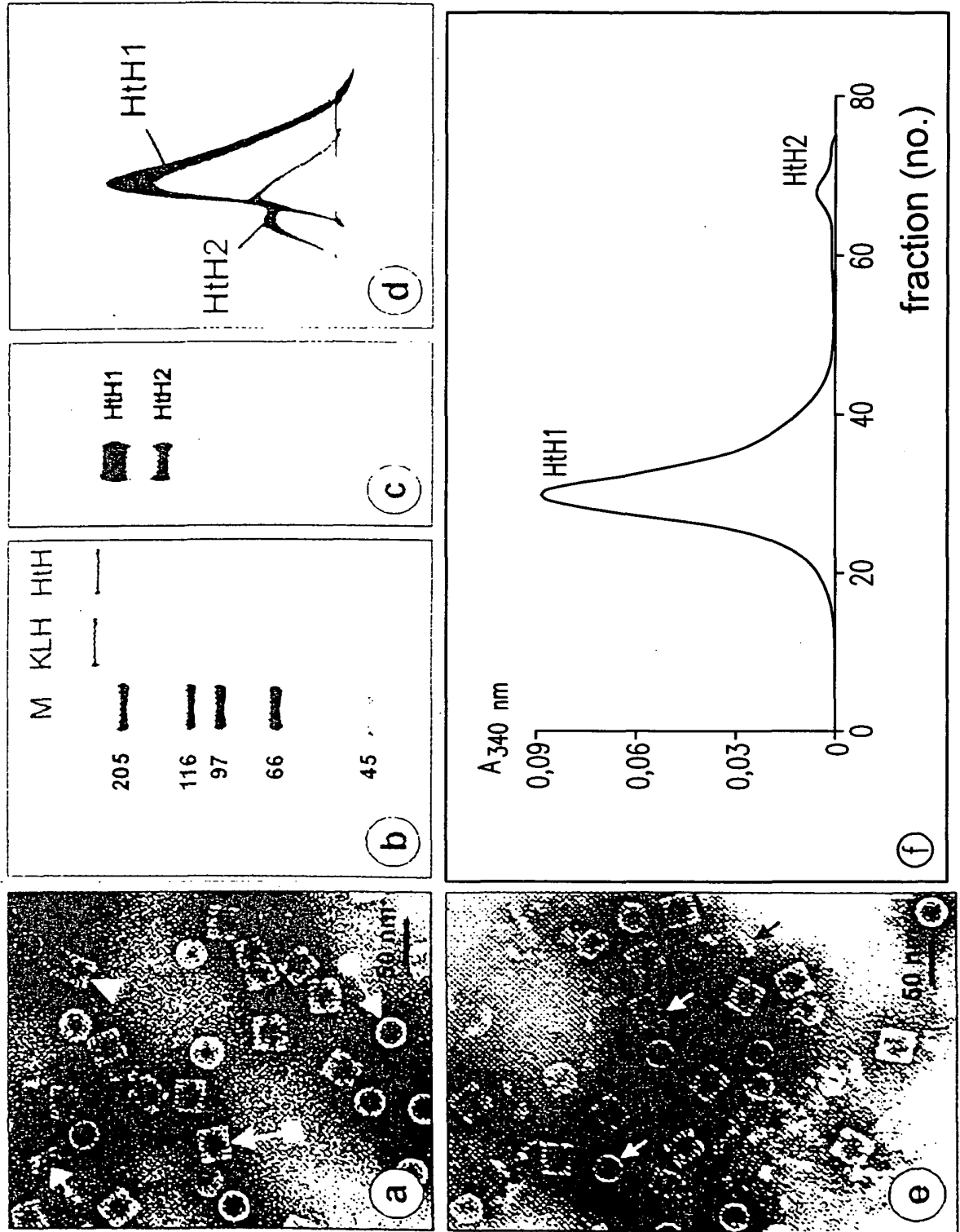


Fig. 1g-m

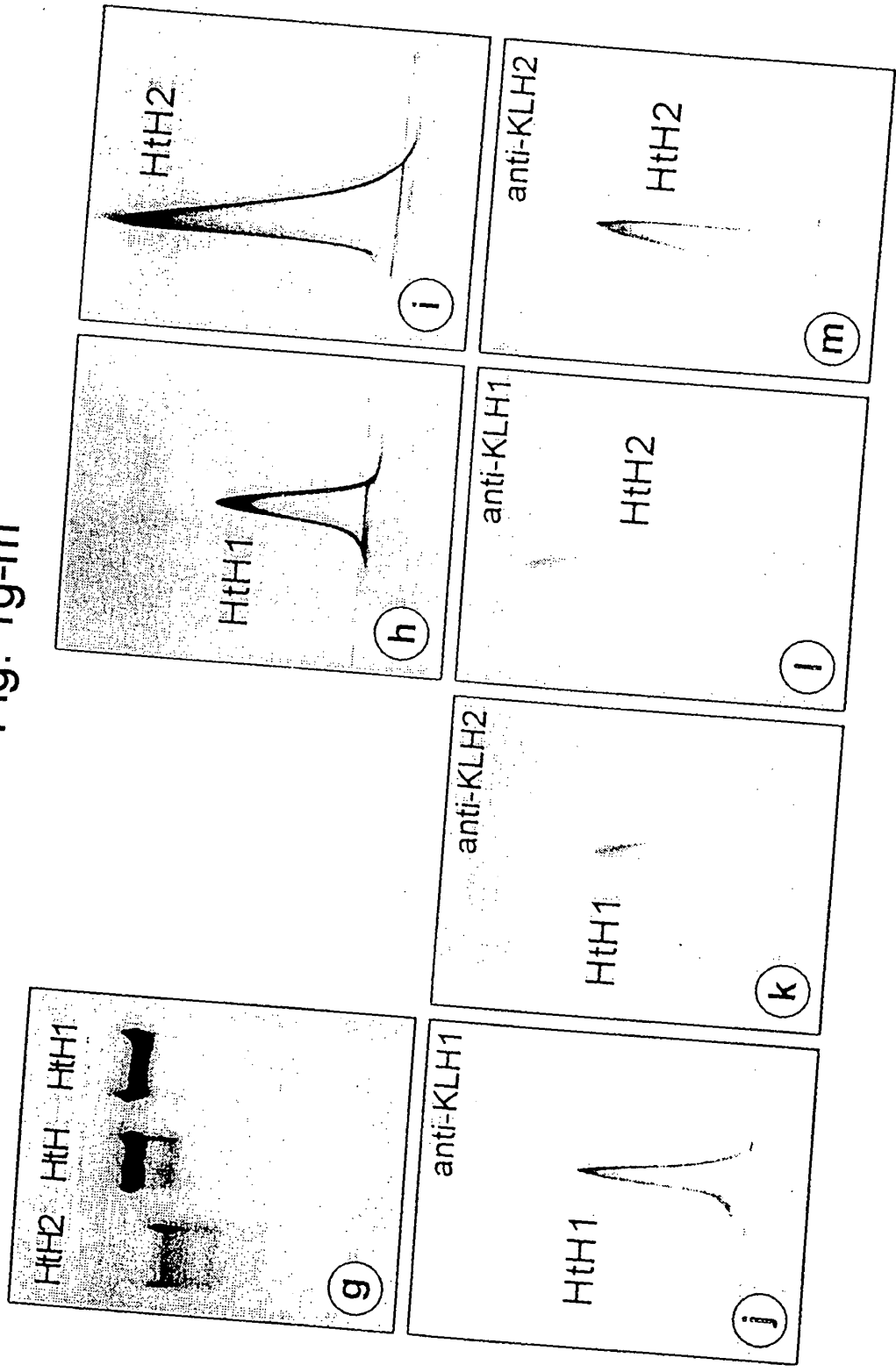


Fig. 2a-h

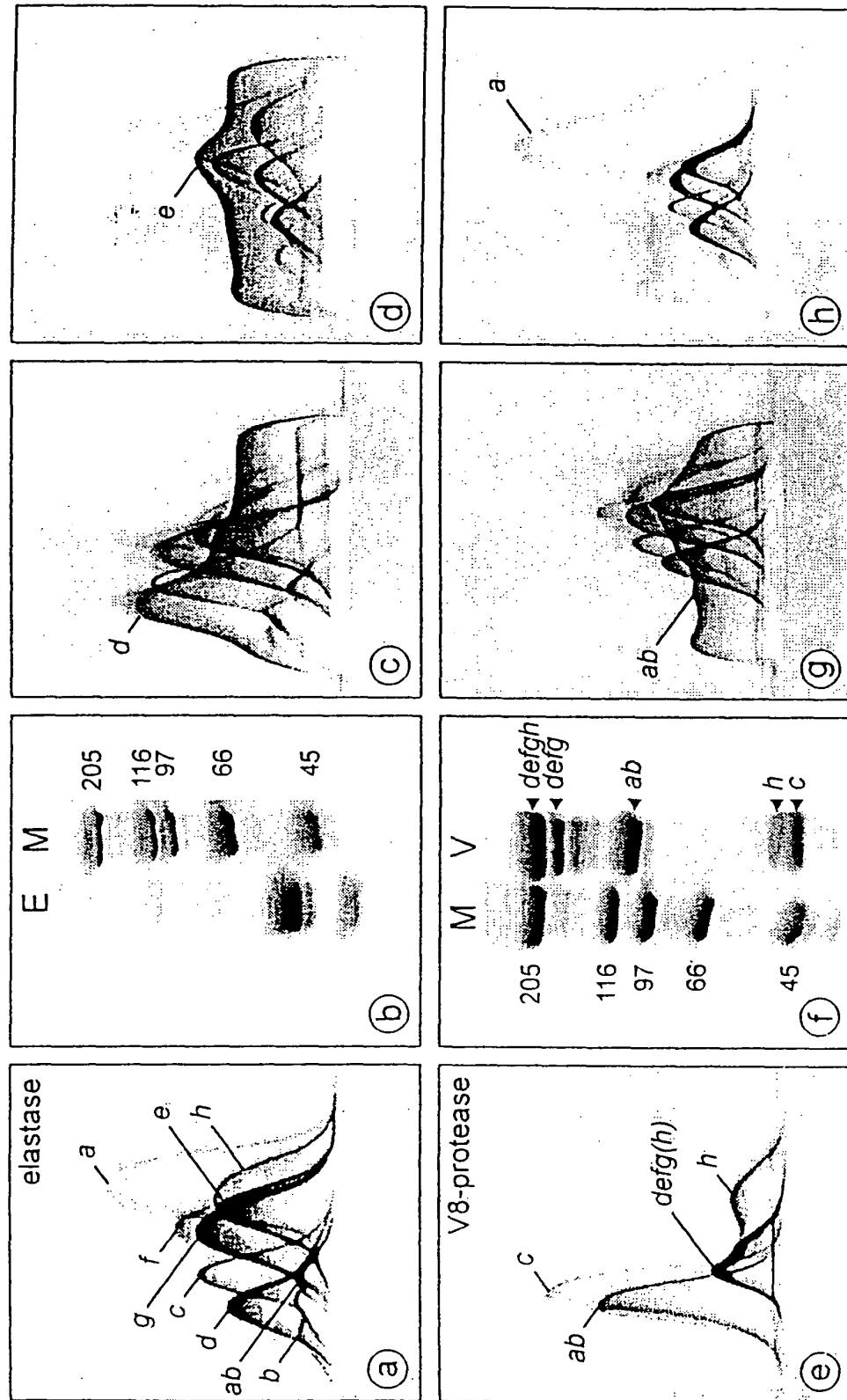
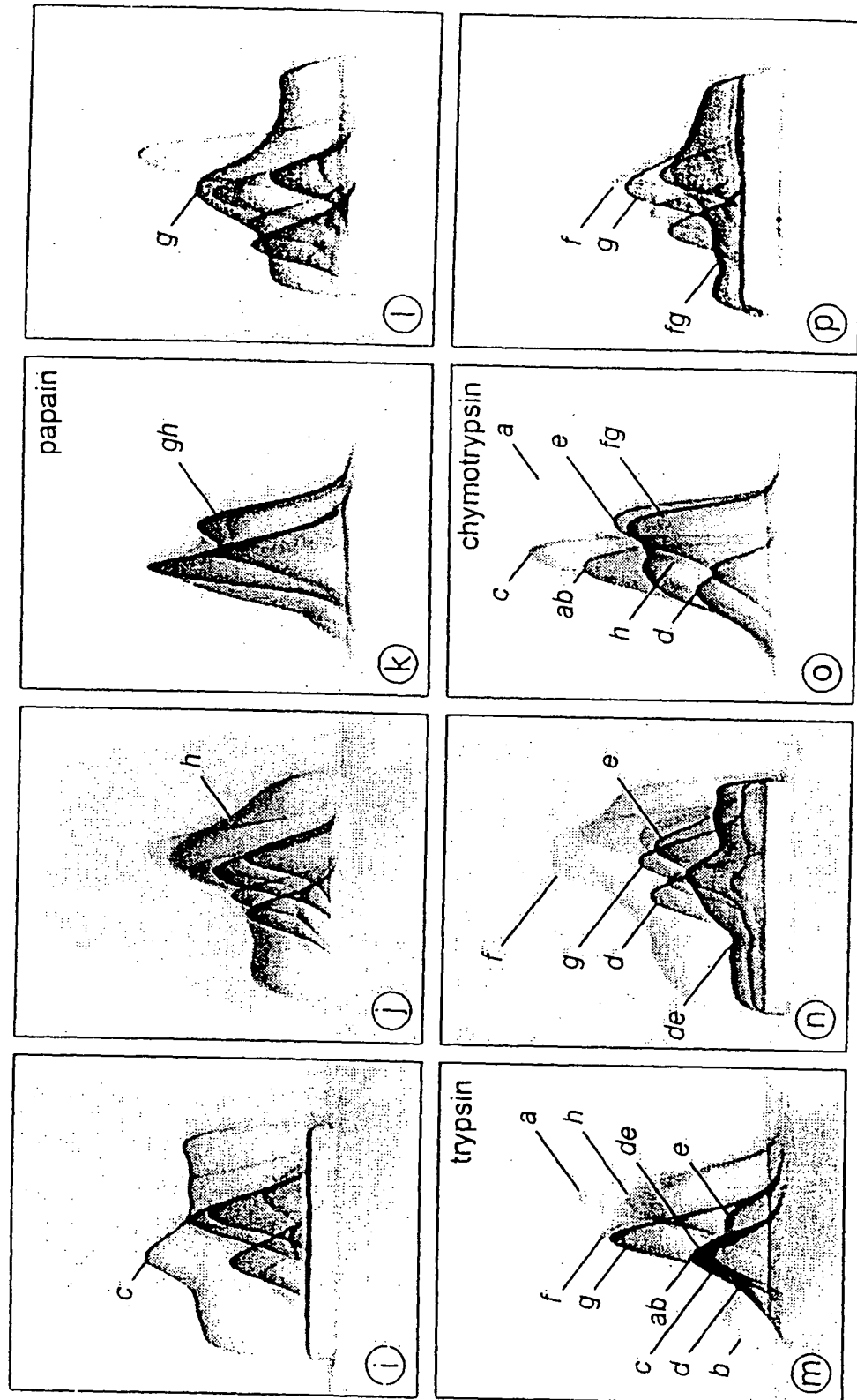
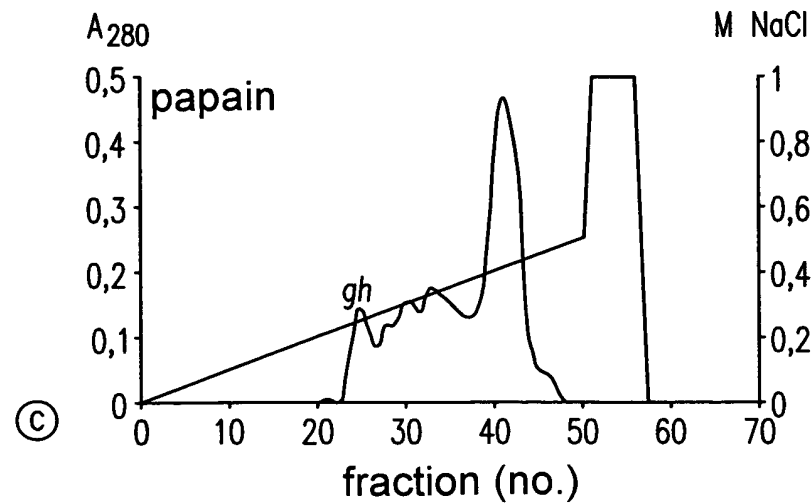
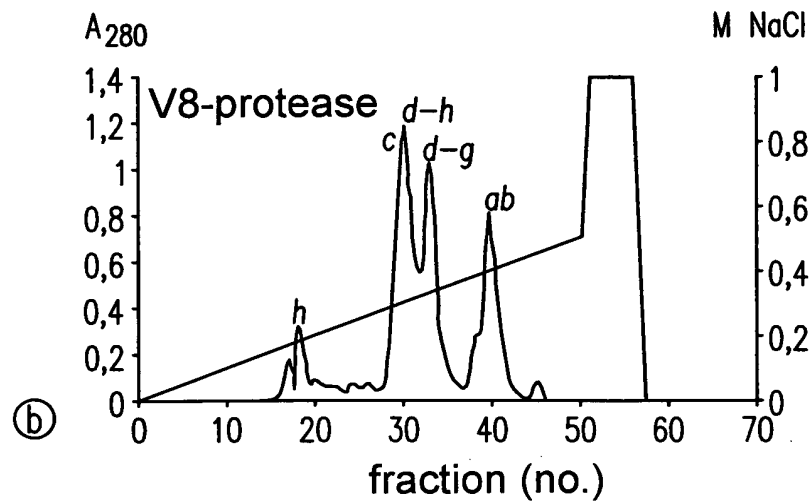
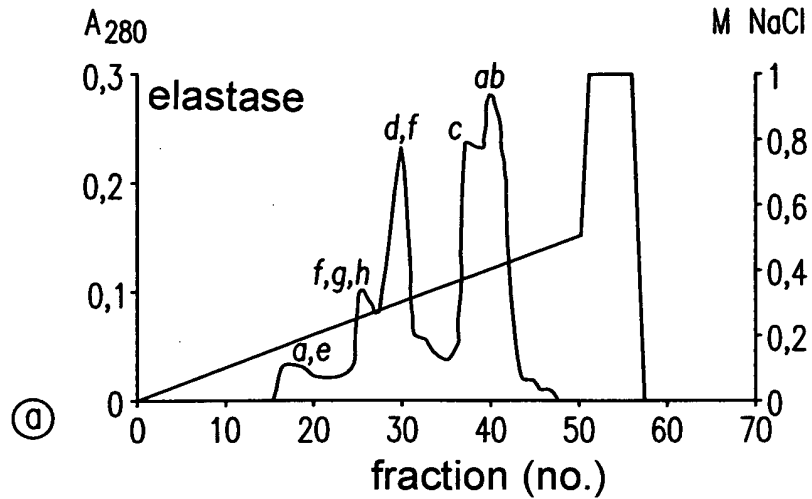


Fig. 2i-p



5/44

Fig. 3a-c



6/44

Fig. 3d-e

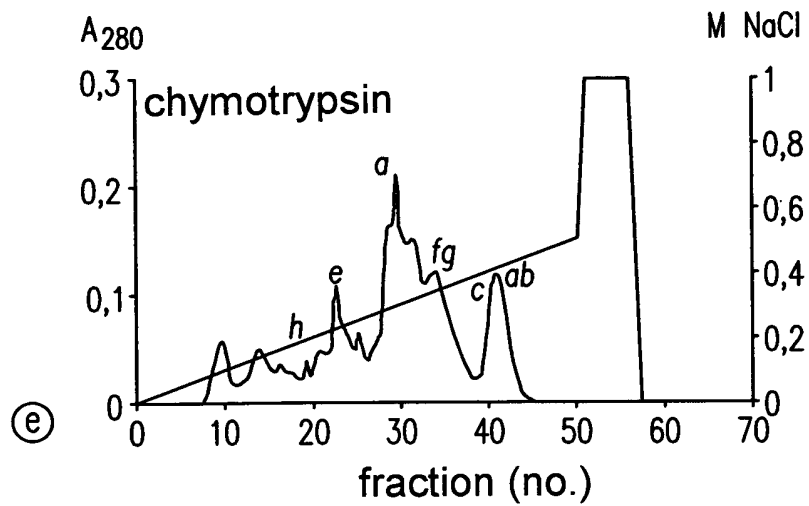
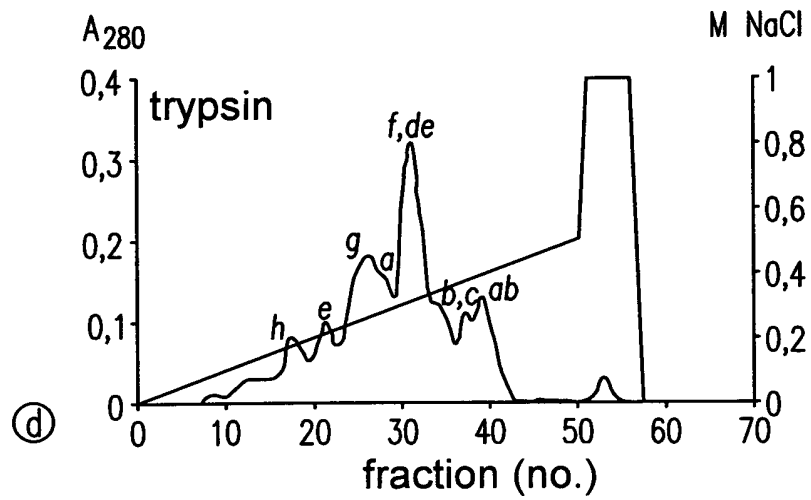


Figure 4**Genomic sequence of the HtH1 gene**

SIGNAL PEPTIDE SEQUENCE 1S-1 (1st part)

GGCTTGTTTCAGTTTCTACTCGTCGCCCTTGTG

INTRON 1S-1/1S-2 (SEQ ID NO:109)

GTAAGTCAACGTCTTTGTTTTAAGTTTGATGCATATCTATCATTGCGTTTTAAAATACCA
TTACAACCAACGTGTCTCTATTGGTCTTCACCTGTTTAACGTATATATTGTTTTTAATGT
GAAAATCTGAGATTATTTTCATTTCCGTCAATATTCGTAAAATACTATACAAATAAAATT
GCTTCAGCCTATTGCATTGGCAGTTTTTCGCAGAATAACGAGGGAAGGCGTACATAAAATA
TAAACCAAGTGTATATTCAAGCATGTTTATAATTTCTTTATAGATTATAACATCATATCAA
AACACCAATCTGGATTTAAACCCGTGAATCCAAAGTATACCAATTAACGGAACCTTTATCA
TGTTTTTATCAAAGTTTTAGATGAGGGTAAAGAAGTCCGAGCTATATTTTGCGATATCAG
CAAAGCCTTCTATCACGTCTTGACACAGGGCTGGTATCTAAACTCGAATCCACAGGAAT
AAATATTTTCAGCCGATAGAGAACAGTCGGTGGCTATCATTGGTCACAAAACAAGTCCAAA
ATCTGCATTAGCCGGTGTTCCTCAAGGCTCTGTCTTGGGGCCACTATTATTTCTCACCTA
TATAAACGATTCAACTAATGGAATATAAAGCAACGTAAACCTCACCGCAGATGAAACACT
AAGTTATAGACAATCCGTTTAAAACCCAGCCACTGCTTAATAATGACTTAGGCCGTCTTT
CAGACTGGGCTAGTAAGCGGCAGGTTAAATTTACCTTGAAAAGACAGAAACCATGGTAT
ATTTCAAAAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGATGATACTGGGATTT
CTAAAGTGTGTGAACAAAAACACATTGGCCTGATCCTACAAGATAACCAGACAGAAACCA
TGTTTTTTTTCAATAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGATGATACTG
GGATTTCTAAAGTGTGTGGTGAACACAAACACCTTGGCCTGATTCTGCAAGATAATGGAA
AATGTTCAGAAACATAAGCAAGTTGATGTGGGGTTTTCTGGGGGTGTGACAACACCGAAA
GACCCTGCAACTAATGTTAGCTCAAAGGGTTTTACACCCGGTCACAAGTGGGGATCGACC
CAGGCACCTTTTTGCCTTTGACAGCTCGCCTTTCAAAAATCTCAATTCGAAAACGAAATC
TAATAATTTTCATGAGCGATACAACCGTTTTTTCATAATGCTGTGGTACCGCATACTGTGGA
AACATCTGTCTACCCATTTGGTAGTCCCCCATAAAATGTATTTATGTTTATAAACACAAT
GTTTATAGGGTTACAGTTAGAAGAAGCATTTCTATTGGCTAATGTACATTGCTTGTTTTT
ACTATTGTGCAAAGGCATATTACAGGTCTTTTAGGAAATTAAATACTGTTTAAATCACAT
ACACTACCGGTAATCCTATTATGCTTATCCTGCCAACATTCTGCCCAAGCAAACGCATGA
AAGTTAAAGCTGAGTGTAATACTGATTGCTGTGTACTTCACAACCAGTGGACTGAAT
ACAACCATGTTTTTTCTTGAAAGTCACAAACATCCAGTCGGTTTTCTAATGTGTAAAGTTT
CTAGTTTTCATAAAGAGCATGACGTAATGGTGAATAGGAGTTATCAATGTTTCTATCTAAT
GACTCCTAGTTCGTTACTTTTTTAATAAAACATCCATGTGTTAATGTTTGGCCACAGAT
ATAACAAGAAAGAAATCGGATAAAATCTACATTTTGACCAATCGGAAGGCTGCCCCCTCC
CTAATCCTAATCATTTTTGTGCCTCAAACATACTCAACCAGACATTTGAACTATGTATA
TATCAGAAATGAAATGGTAACAATAAACTTGTATGTTGACCAGACAGAATTAGGGTGAATC
TGAATACCAACTATTGTCACATATGAATATGGATAAGCTCTGCGCGTGCGTGCGGGCGGT
GTAGTGCGTGTGTCTGTGTGTGTGTGTGTGTGCGTTTTGTGTGTGTGTCTGCGTGCGTG
TGTGTGCGCGTGTGTGCGTGTGTGTGTGTGTGTCAGTGTGCCGAGTGTGTGTGTGTGTG
TGCACAGACATGTGGTTGAGACACACTTGATTTCAGTGCAGGATTATGTCCTTCAACCGAG
TGTAAGTCTTTAAGTGTGCCTGGAAACAAAAAAGTGCCTTGGGTTCATCGCCTCTGTAGC
AAGCTTGGACGCGTCACGCAGCTCTGATACCACGTATTGGCACCATTGTTTCATCGGTCTC
ACGCGAATATTATGCTATGTGTGGCGTATCATACCATAGGTTGGGAACGTTTCAATACTG
TACCGAGCTTGGGCGTGTCACAAAGCTATGATAAGATGACAACACGTCTTGGCATCTTGT
TTCCTCGGTATCACGCGCTGTTATGCTATGTGTGGCTATCACACCTTAGGTTGGGAAAGT

TTCCACATTTTCCAGCCTCGTACATGTTTCCTTTTGTTCCTTTTTCCTTAGTTATCAGCATAC
CGTATATTCTATATTTAATGAGCATTGTATTTTCTACAG

SIGNAL PEPTIDE SEQUENCE 1S-2(2nd part)

GTGGGGGCTGGAGCAG

INTRON 1S-2/1A-1 (SEQ ID NO:110)

GTGAGTTTCTTAACATTGTCATGGTACATGGATATACGCTCAGTGGGAAAGCAGGATATC
CCCTTGGTTCAAGTATTCACTTGTACGCCAAGTGTTTCGATTCCCAACATGGAATACTGT
CATATAGTAAATTGATACTACTTACATTTAATTCTCCACTAAACGTCAACGTCCTTTA
CTTCATGGCCCATGGTCCGTATTAGTGAGTGAGTGAGTCAGGGCATAAGTATTTAACG
TCAAATCAGCAATATTTTCAGCCATATTGTGACAAGAATTGAATATAAATAATTATACTTA
TAATGCTTATAAATATAAATTATATAAATACCTATAACTATAAATTAGTTATACTAGTAT
TTATCAAAACATATTTGCCACGACACTGCACGCCGATACTTCAAGTGTCTTCACCTCAAG
CGTGTAACCTCCTCATACTCTGTAATAAGTATGTACACTAAGTGAGTGCTATCATCTCCAT
GCTTCATTAGTTTCGTCAGATGCGTGTATCCATACGAGTACATTCAGATTATGGGATCCA
GAGCTTTCTTATCTCAAGTATTTCCGATTGTAAAGCCATACTACTTCCCCAATGACTGAC
GAGACAGATGGCAACCGTTCTTTCCTCCTGACTAGGTGAGTGCCACTGATAAATCATTAT
GCCTTTAACATTAGGAATGTTAGCAGTGCACATGTTTCAGAATTGCGACCTTATGGTTGT
AAAGATTACAACTTTACAACCTACTTGAGACAGGTTCCATATGTCGTATCTGAAATAGT
GTGAAGGTATCTGATTTCGATGCAATACACAGACATATAAACATATTGTCGCCCTGCTATT
CCGGAAAGGTCATTTTGTATGTAACGTTTCCTTAATGGACACAAACGGAATTATTAGTTAA
ACATACTCAACAAAACCTATGTTATTTTGCAATGGGTAGCACCGAAATCTACCGACAGTGG
TTCGTAAAGTAGAACATTCTGACATAAAGAAAAATCATTTGGCTTTAAATATATGCAAGT
TACTTGTCTCTAACAACCGATTTTATACACATTTTCAGAGAACGGGGAATCCGCGATGACA
ATATCAACGAGTATATACAGAATATATAATTAAAAACGATGAGTGCCTGGCAAGGGAAAG
AGCGAGATTTGCCAAACAGGGGGGTGGTGTGAGCTTGAATCGTGGAGAAACGTAGATTG
AAAGACAAGATGACATCTAATGATCCGAAAATCAAACACAGGATTAAGTGGGATGCAGAA
GAATGAATATCTCAAGCATAACATGCAACACTTCATGAATGCATCTCAAACATTTTCGTCA
GATCGGATGCATGAAGATTTGTAAAGCAATGGTTTAAATTGTCCCTAAACGTTTAGTTGG
AGATGTATGAGGCTAGGCTGTATGTTGAACGAAACCATTTAACATTGTTGTTTCATGATTA
TTTAATATTTTTTTCATTTTATAGATGTACAATAAAATTGGAAACTAAACATTTCCCTTTA
TTGTTTTGTATTTACCTGTTTCATGGGTATGTTTTGAAAGATCGTGATATTTAGTTGGCAT
TCACAAGTTGGAAAAAGGTCCTCAGTTTGATTTCAAGTTTATGTAACCTCTTTATCTGA
CGCTCCAAAATATGTATAGCCTTGTTTCATCTGTCGGTATGTGGATATTCCTACTTCAGGG
TAGGGTAGCATTAATACTTACAAAACATAACGTGTACCAGATTTTCAGTCACCTCAGAGAT
GATAATGCATGTCGATATGATAGGTCAAAACTTTCGATATCAATCACAATGAACCTATGG
ACCTGAATCGGAATGATACGTTTACCTTTAGAAACAATTCACAAATATGACTGTCACCC
TTTCAGGTAATAATGTTTGACGGACTACGATAGTGCTGAACAGCAGGAGAGGCAACATGG
TTCGATTGTGAGACAGGTTTAGTGTATTTGTTTGCGAATTTAAGGTTCTGAATCACAATA
GACACGGTTCAGTTAATGGATAAACCAATCATTAGATAGATAGAGATTAGTCGCGATATT
GCTGGGATAAAGCTTAGTGGGACGTTAAGTCCCATCTCAATCTCTCTCATTTTTTCCAAA
ACAGTTTTTAATTCAGGCTCATGACAAGGTCGTACTGTTGCAAAGGATTCTACTTCAAGCA
GAGATGTCTCATGAATACAGTACAGGGTTTTTGAAGTTTATCCAGTGCAGCGCTGGCACC
ATCTCTGCATGCGAATTATACCATCCATGCCGCTCTAGGCTATTTGTATTAAGTCTGTAG
AATTAAATTCGCGAGTTGCAAATACTGCTCACCATTATCTGCCTCAACCCAGTTTGGGTA
CATGCGATTTACACAATATTATGTATAATGTTTCGCTTTTTCGAAAACAAAACACCTAAATT
CATCCAAAGTTTTTGGGAGATTTTATTCGAGAAATCAACCTGAGATGTTGAATCGGGAGCT
GCGCTTATTCAATGGTGGACTCGGAAGGGAAGTAACCGCTGATGAGGCAAAACAATAACG
CAAACATATGGAAGTGGAACCTCTTTGAACCAGTATTATGTTGTGTGGACATGTATGTGT
TAATTTGACCATTTCGAACAACCTTACTATTCTATTTCATAATGTGTTTAGATTTACATTTG
AATTAAAAGAGATGAGTTTAAGATATTAATATTTTCCTTTTATAGTCTGTTCGTGATTGTA

GGGCAATATTTATGTATGTTTCGTTTCATTTTTTCATTTATCATTTGGAAAGGTATATCATAA
 GATTATTATTATCATTTCTTGAAGTAATGTATACATATATATATGTCTTGAGTAGCTTATT
 TTCAATTTATTATCATCCGTCATCCAATTTTATTTTACGAAAGTATAAGAAATAACGAGA
 GAGAGAGAGAGAGAGAGAGAGAAAAGACAGAAATGAAGTTAGGAGATATNAGTTATCAAGAA
 AACAAACAGTTTGAATTTTTTGTTTAGACAAGATATCATATCAATAACCTCGCACTATTAC
 GGGAAATAGGCGGGCGTTCCATATGCACAATGAATCGTCAGTTAAAATCAACATTAAACTT
 AAAATACTCCTCATATTTAAAGTTGATCTACCTCTTGTATTATTGTAGACTATTAGACAG
 AAGTCGACAGTGACACCAGCAACCAGATATCATACCCAGACTTAAAAAGCTGTTTCCTTG
 ATGTTTCAATTTATTTCCATTTCCATTATTTCCCTTTATTGGTTTCCATTTATCAAACCTT
 ACCATCTGCACCAGTGGGAGATTGATATGTTGTATTTATTTATATTTCTTGTACTACAAT
 ATCAAGAATGTATAGGAGCTATTCCTTGTTCCATAAACCGGATAGATCCATAATTTCCAT
 TTTGGGATAAATGGAACTAAACACAACCTTTTACAGTAAACACGAGTGAGCAAGTTGAGT
 TTTACGCCGTTTTTTAGTAGTATTCAGCAATATCGCGGCGGGGACACCAGAAATGGGCT
 TCACACAGTGAATGCATGTGGGGATTTCGAACCCGGGTCTTCGGCGTGACGAGTGAACGCT
 TTAGCCACTAGGCTACCCACCCGCTATTTATAGTTAAGACGAATACTTTTCTCAAGCCT
 CAAATATGTCCATTCTAGAGAGACTGAATCTGATCCTGAATCTGCGGACCGGTCTTGAAT
 ATCATCCCACTAACTCATTGTACAAAGTACCTGTAGATTGTCAGTTCAAAGACAGATTTT
 ACAACCCCTATTATATTTTGTCTGCTCATTAAGATATTCAGACTCACTCAAACCTGCTAAA
 TGATTTTAAATCCTACTTTGAGATGTTTTAACTTTTATTTCGATGCATTTTTGCGTTCTGCG
 TCCTGTATAAAGGTAAAGCAGGTAAACTAACCTAACCTGTTGATTTATTTTCATAGTTTTG
 CGATCAGATTGAAACCGGAATGCACAGTGAAGTGTGGCATAACATCTTCCACAGAGATAC
 TGGATACTAGGTGGTACAACCGCATTGGCTTTGTGAAAGGATATTAGTGTTTTATGAGAC
 TGACTCATGTTTCAATGCTTAGAGCGGAATGATCTCGGTCTTCATGAAAAATATTGTGTT
 GAAGTAACCCCCCAGTCCCTAACAGAACGTGGGGAAAGCAGATGGATATGCCAAGACATC
 TTCGCATGGTGTGAAGATGATCGTTACAACATCTGCAGAAAAAGTTATTTCTGTGAAGAA
 TATGCCAAAGCATCACTGTGAGTGTTTTGAAGATGTGATATGGCAACACGCAGCGTGTA
 TTATGCTTTGTGTGTATTTCTGAAGATCCGTATGAGCATGGCGCCAAACTATCAGTTAAA
 TGGCTATGCGAAGATCTTCCCGAGATGGTAAACACATATTTTGGCCATTTTCTTTGTAAG
 TGGGCGACACAGAAGATCCCCCTGATTGTGTGGATGAGGACACAAAAACGGGTCCCCCTT
 CCTTTGCTGATGCTAATGACGCCCTGGAAACATTGAAAGACTTCTTCTCCAGCAAGCAAG
 CCACCAACCACAAGTTGTATAAATCGCTTGCGGACTTGAATACGGCAGTTGGACAGATAC
 ATACAGCCAGAGAGGGCCGAACATAAACATCTAAACATGGAAAACTGTAAAGACAGGCT
 TTGTTGTACGACGTACGTAAATTCATTGAATGTTTGAAGAGGTAGAAAATTATTAAATCT
 TTGAAACCTCGCTCTGTTTGTATTGTATTTGCCCCACATTTGCAAATGGTATCCAAAAA
 GGGCAGACACATTTGTTTTAATCTTAGCCAGGTTCAATTTAGCCTTGCGCCCAGACTCAT
 TGTATCTGGTGAAGGCTATAGGTGGCCACGTCTTCTAAGATGCTATGCTATTCTTACCAG
 AATCCAATGTAAAGAGTTCAAACGCATGGTTTCGCTTTGATTGTGATTCTTTCTTAGCACC
 TCTCTCTACCCAGAGTTCACCTGCACTGCTCCTGACTCACAATAAGCTGACGTGCTGTC
 ATATATGTGCAACATTGTATACGTTGGCGTTAAGCCCAACTCACTTCCGCTGTCTTTTGG
 CAG

DOMAIN 1A-1 (1st part of domain a)

ACAACGTCGTCAGAAAGGACGTGAGTCACCTCACAGTTGACGAGGTGCAAGCTCTTCACG
 GCGCCCTCCATGACGTCACTGCATCTACAGGGCCTCTGAGTTTGAAGACATAACATCTT
 ACCATGCCGCACCAGCGTCGTGTGACTACAAGGGACGGAAGATCGCCTGCTGTGTCCACG
 GTATGCCCAGTTTCCCCCTTCTGGCACAGGGCATATGTCGTCCAAGCCGAGCGGGCACTGT
 TGTCCAAACGGAAGACTGTGCGAATGCCTTACTGGGACTGGACGCAAACGCTGACTCACT
 TACCATCTCTTGTGACTGAACCCATCTACATTGACAGTAAAGGTGGAAAG

INTRON 1A-1/1A-2 (SEQ ID NO:111)

GTAAC TACAAACGTCGTC CCATTCATACAGGAGAAATATACAATTGTGTTGTAAGAGCGG
 TATACTGTTTGCCAACGTGTGTAATTGAAACGTTGATGATGGTGTCTTTGTATTTCAATTT

10 / 44

GTATGCACTTAGACATGATCAATGTTTCTGATGTGTCAAGGATGTTTCGGTGTGTCACTTT
CAAAAGATCAAATTCATATGACGTACACAGAGCAAGAACCAACAGTAAGAAGTCTGTATG
ACTTCGCTCTTAAAGCAATGGAAAAATATTTTCACTTAACACCTAGCCCATAATCACGC
ATATTAGATTATTTCAAGCGATGTCAACATGTTTTTAATATCAATCTCATGGTTCTGATAT
TACCGGAGACATGCAACAGGCTGCCATTATAGCCAGGAAATCTTATGAATATGTGCATAT
TTTTTCTTTGATTCTGTATGACGAGAAATATTCGGAGGCAAAGATTGTGTTTTCAGAACA
GAATCAGGGTATCAGTGACATCGTCACTGCATGGCTACAATATTGCTGATGTGACTGTTT
CTCCAAGGATTTTCATCTCACTGTCTGTACTTTGAATCTACAAATTCGTATTAAAGTTAT
GACAATTTTACCCCTGCCTATTTGTAAACGAAATATAACATGAGTGTTTATGCTGACAG

DOMAIN 1A-2 (2nd part of domain a)

GCTCAAACCAACTACTGGTACCGCGGCGAGATAGCGTTCATCAATAAGAAGACTGCGCGA
GCTGTAGATGATCGCCTATTCGAGAAGGTGGAGCCTGGTCACTACACACATCTTATGGAG
ACTGTCCTCGACGCTCTCGAACAGGACGAATTCTGTAAATTTGAAATCCAGTTCGAGTTG
GCTCATAATGCTATCCATTACTTGGTTGGCGGTAAATTTGA

INTRON 1A-2/1A-3 (SEQ ID NO:112)

GTAAGTTTGGTTTACAGTTTCATTATAAAAACATAGCAGTTTTAAGTTTAGGGGCAGATT
CTAATCTCTAATATTCCTTTCAACTCACTTTATTGGTGCCTTCTTGGAGTGACATTTAGA
AACTAAGACAAGAGGAAGATGAACAATGTTTGTAGGGATAGACAGCTTGGATGCAATTTT
GGACCAGATTCTAACAGCGTCATGAAGCAAGTGATACACAACGTTATCAATAACGAGAAT
ATACACATAGATGGTTTGAAGTTTATAAATGAACTATTAACGGCATTGTGGTTATAGACAG
TGAGGAAGACGCCAGATAGACAAAGGGTAGGGGCCTTGGTTAGATAATGAGAAGTTGAAG
AGGTGTAATAACTTAAATCTCTCTTGACTATTGATTGTGTCTAAGAGTTTTCTTATCTTA
CAGTCGGCCAGTTGGGTCAAAGATGGTGTGATTTCGGATGTGCTTTGTGTGTTCTGCGATG
GCTGATTTAGAGTCAGTTTACTTCAGATGAATGAAGTTCCCCGATTCTTATGTTTAAGTT
TGTTTACCTACGCATGAAGACATCACCAGCAGGGTTCGTCTTTATTTCTAGTAGCTTATT
TACAGCAAGCTTGTAACGTATGCTGAATTGCTGTGCCTCTGTAGAACACAGCATCTATGT
TTGCTTGCTTCTTTAGTAGACTGCGGATGTGATGGTTGGTTACCTGGTATGCTGACGAAA
GAATTGTTGACGTGGTGGTTTGCCTTGATGGGTTCGTTGACTTGGTTTGTGGATACTGA
TTAAGGTGACTCTGCTGGGAGGCTTGGATTCTGGGGCCGGTGTTCTTTGCTCTCCTGTCT
AGGGTGGCGATTATTTCCCAACCCACTTGTTCCATTACACTCAAACCTGCTATCAATTT
ACAG

DOMAIN 1A-3 (3rd part of domain a)

ATATTCAATGTCAAACCTTGAATACACCTCCTACGACCCCATCTTCTTCCCTCCACCACTC
CAACGTTGACCGCTCTTCGCCATCTGGCAGCGTCTTCAGGAACTGCGAGGAAAGAATCC
CAATGCAATGGACTGTGCACATGAACTCGCTCACCAGCAACTCCAACCCTTCAACAGGGA
CAGCAATCCAGTCCAGCTCACAAAGGACCACTCGACACCTGCTGACCTCTTTGATTACAA
ACAACTTGATACAG

INTRON 1A-3/1A-4 (SEQ ID NO:113)

GTGAGACATTATTACACTTCTATTTAGTAGTGGGGGCGGGATAGCTCAGGTGGTAGAGCG
TCGGCCTTCAGCTTCTAGTCTCGCCCAAGAGCGCGCTGGCTAAAGGGCCGGAGTTAGA
TTCCCGCGGGCGGCAGGCAATATCTCCGAAGGGGAGAACAGTTCTCCAGTCGGTGAAATT
GGGGTGCAATGTTGTACCACTGAAATGCGTGCAGCACCACCAACCATCCAAATACCAGCCTTG
CCGCGCTGGTCTGACTACATAGTACCACCCGGATTCAACCGGGCTATATAGGTTCTCCTC
CAGCAGTAAATCTGACAGTCGCCATATAGCTGGGATATTGCTGAGTGCGACGTTAAGCCC
CAACTCACTCACTTTATATTTAGTATTCTATTTAGTATCGACGCATGACCATGTGTGGTG
GTCTACTCATCTCAACACGACCGATTAAACGTTAAGAGCTGCCAACATGATTCTCTTTCTC

TCTTTAGCCTCTTTATGCCAAAAGCTATATATTAATGTAGGACCCTACATATATTATTTTC
CAG

DOMAIN 1A-4 (4th part of domain a)

CTACGACAGCTTAAACCTGAATGGAATGACGCCAGAACAGCTGAAAACAGAACTAGACGA
ACGCCACTCCAAAGAACGTGCGTTTGCAGCTTCCGACTCAGTGGCTTTGGGGGTTCTGC
CAACGTTGTTGTCTATGCATGTGTCCCTGATGATGATCCACGCAGTGATGACTACTGCGA
GAAAGCAGGCGACTTCTTCATTCTTGGGGGTCAAAGCGAAATGCCGTGGAGATTCTACAG
ACCTTCTTCTATGATGTAAGTGAAGCGGTACATCACCTTGGAGTCCCGCTAAGTGGCCA
CTACTATGTGAAAACAGAACTCTTCAGCGTGAATGGCACAGCACTTTACCTGATCTTCT
TCCTCAACCAACTGTTGCCTACCGACCTGGGAAAGGTCACCTTGACC

INTRON 1A-4/1B (SEQ ID NO:114)

GTAAGTTGATTGTCTTAATATTGTTTTAATTTTTGCAGAAATTTGATTTTAAATTGTGTA
ATAACAGTACACATTTTTACGCAACAGCAGTCATTATTGTGTGTGAAGATGTCAAACCAG
AAAGGTTTCAATCGTGAAAACAAAAACAATTCTCTATCTGTATACCCCTCAATACCAGTA
TGATCACAAATCTAGGAAATATTACAATACTGCTTCATAGAGTAACTGCTGTTTGTGGCA
GAGCTGGATACGAAGTTTCTGATAGTTCACAGCTACATGATAGTAAATGAACCTGTACAC
ATCAACGGTTGATCATGAAAATTTTTGTATGTGTGAAAGTGCTACCTGTATTAGTGAACGT
GCTACCTGTATAACTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATGCTGAAA
GTGCTACCTGTATTAGTGAACGTGCTACCTGTATAACTGAAAGTGCTACCTGTATGACTG
AAAGTGCTACCTGTATTAGTGAAGTGCTACCTGTATGAGTGAACGTGCTACCTGTATAA
CTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATTAGTGAAGTGCTGCCTGTA
TTAGTGAAGTGCTACCTGTATGACTGAGCGTGTTACCTGTATGACTGAACGTGCTACCT
GTATTAGTGAAGTGTAATCTGTATGAGTGAAGTGCTACCTGTATTAGTGAAGTGCTA
CTTGTATTAGTGAAGTGCTACATGTATGACTGAAAGTGCTACATGTATGAATGAGAGTG
CTACCTGTGTGACTGAAAGTGCTACCTGTATTAGTGAAGTGCTACCTGTATGACTGAAC
GTGCTACCTGTATTAGTGAAGTGCTACTGGTACCAACTGGATGTTCTCACTTCTTTGGC
GAATATCTGGGCTCAAACAGTTTTTTCAGTATCATAGTCGTATCAGTTTGATTTGTATGT
GCAGTGAATCATTTTTCGTCAAATAATCAAACCTGGTGTTGAACTGGCGTTCACGTTTTA
TGGTTGTAAAACAAATTCTGTAAGTAAAGATATTTTAGGGATATCTGTATGACATGAACT
GAATTGCTTAAGGTTAGCATGCCATGACAAATTGCTGAATGTCTGAGGATTGGTGGAGCA
ATAAATCATTATTAAGACAAAAATCAGAAACGTCCATTTTCACTTTTAAACAGTGTATCTG
TCTGAATGCCCCCTACTTTTTGGAAGAGTATATATGAATTATCGGCAATATAAAACGTTA
AATGGCAAATGTGCGGCATATGTCAGGACATTATTACCGCAGTTTATAGTCATATTTACC
GGGTCTAGGACAATTGTCACCCCGACAATTGCCACCCGGACAATTGCCACCCAAAAATAA
AATATACGTAAACAGAAAACAAATATTGCTTTCAGCCTTTATTGAGTTAGATAATGACAT
TTATGTTGATAAATATGTCGTTTGATAATAATAACAATAATATAATATTACAATACT
GCAATAGTACTATCAGTACTTATCATTTTTATCACAGATTATATATAGATTCTAGAGTCCG
ATGTTGTAGGCAACACTTCGTCGGTAGGCCGTTAGGTAGTTATCATTAGGGCTGAGTATT
GCGCCAAATTTTCGTATTGCTATATACTGCGATACACGGTTACCTGTTTTGCAATACGTAA
ACTTAGGCAAATATGACAGTTTTTCCATGATTATTTTCACGTTTCAATGCTTAAATGGT
CTTATCTGTTATCTCCTTGAAGGTTTAAATAAATAACAATAAACATAAATCATTATTGAA
AATTAATGAACAAAAGTAAAGCGCTTCTCAGTTACCTTAACCTAACTTATTTATGAATGG
GATTACTATCCAAGAATGTGAAATTCACAAACACCTTGGGATAACACTGCAAAACGACTG
TTCATGGGACGGACATGAAAAAGGTGAGTCCCATGTTAAACTGTTGAGAAAAGTTTCTAT
ACTGTTTTGTCCCGAAAAAGGCTAAAGACCATGTACTAATCAATTATTTCTATCTATTTTCG
ATTACTGTTCTCATATTTGGGACAACCTGTGCAGATCGGTAGCATCCAAGCTCGTCTAAAT
CGGTTTGATAAACCTTGTCAAATAACATGTTGTCTCAACATCCAAGCTCACCTAAACCTT
GTCAATACCTGCATCTGAACAAATGTATATTTAAGACGATAGCATCCAAGCTCATCTTTA
AAATGAATATTTTCTCTTTTCTACCAAACATTATTTGGTTGACAGTTGTCTCCCTAT
TATAGTAAAAAGAACTGGGTGGCAATTGTCCTAGGTGGCAATTGTCCGGATGGCAATTGT

CCGGGTGGCAATTGTCCGGGTGGCAGTTGTCCAGGTGGCTATTGTCCTGTTCCCATATTT
ACGTATCCCATTTTTCTGCTCTGTAATTTTAAATAAACTCACCTGCCTAAGGTAAGACGAC
ATGTGTCACGTGAACATCGTTTGGGGGCAAGGGCGGAATCCCTTCGTTGAAAGTAAATGA
ATACTGTACATAGAGATGCGTATCTTGAACCTCTTATTAGCTTTGATATTGTGCTTAATA
TTACATGAATGTATTTCAATATGTAATTATGTGTTCAAATGAATGGTTGACTTGAATGGT
TTTATTGCTTTATATGCTACATCAACATGTGTGTTTCTTTTCATTTACAG

DOMAIN 1B

CACCTGTGCATCATCGCCACGATGACGATCTTATTGTTGAAAAAATATAGATCATTTGA
CTCGTGAAGAGGAATACGAGCTAAGGATGGCTCTGGAGAGATTCCAGGCCGACACATCCG
TTGATGGGTACCAGGCTACAGTAGAGTACCATGGCCTTCCTGCTCGTTGTCCACGACCAG
ATGCAAAAGTCAGGTTTCGCCTGTTGTATGCATGGCATGGCATCCTTCCCTCACTGGCACC
GGCTGTTTCGTTACCCAGGTGGAAGATGCTCTTGTACGGCGTGGATCGCCTATCGGTGTTT
CTTATTGGGACTGGACAAAACCTATGACTCACCTTCAGACTTGGCATCAAATGAGACGT
ACGTAGACCCGTATGGACATACACATCATAATCCATTCTTCAATGCAAAATATATCTTTTG
AGGAGGGACACCATCACACGAGCAGGATGATAGATTGAAACTGTTTGCCCCAGTCGCTT
TTGGGGAGCATTCCCATCTGTTTGATGGAATCCTGTACGCATTTGAGCAGGAAGATTTCT
GCGACTTTGAGATTGAGTTTGAGTTAGTCCATAATCTATTTCATGCGTGGATAGGCGGTT
CCGAAGATTACTCCATGGCCACCTGCAATTACACAGCCTTTGACCCCATTTTCTACCTTC
ATCATTCCAATGTCGATCGTCTATGGGCAATCTGGCAAGCTCTTCAAATCAGGAGACACA
AGCCATATCAAGCCCACTGTGCACAGTCTGTGGAACAGTTGCCAATGAAGCCATTTGCTT
TCCCATCACCTCTTAACAACAACGAGAAGACACATAGTCATTTCAGTCCCGACTGACATTT
ATGACTACGAGGAAGTGCTGCACTACAGCTACGATGATCTAACGTTTGGTGCGGATGAACC
TTGAAGAAATAGAAGAAGCTATACATCTCAGACAACAGCATGAACGAGTCTTCGCGGGAT
TTCTCCTTGCTGGAATAGGAACATCTGCACTTGTTGACATTTTCATAAATAAACCGGGGA
ACCAACCACTCAAAGCTGGAGATATTGCCATTCTTGGTGCTGCCAAGGAAATGCCTTGGG
CGTTTGACCGCTTGTATAAGGTCGAAATAACTGACTCATTGAAGACACTTTCTCTCGATG
TCGATGGAGATTATGAAGTCACTTTTAAAATTCATGATATGCACGGAAACGCTCTTGATA
CGGACCTGATTCCACACGCAGCAGTTGTTTCTGAGCCAGCTCACC

INTRON 1B/1C (SEQ ID NO:115)

GTAAGTAAATTTACAAAATTTGGTGTTCTCTAACTATCCTAAGTATTCAATCGTTAGCGT
GTACCTATCTGCATAATGCAATACCCTGACTCCATATAAGTATAGTATATTTACTCTGGT
CGAAAACAAACAAATTGAAAACAAGAGTGGACGTGCTGTTATGATTTCTTTTTTCATTCTT
GGTTCGTTGTGTAATGCCACAGCCAGCAATTCCAGATATATAGCGACGGTCTATGAATAC
TCCAGTCTGGACCAGACAATCGTGTGGAATGGTTTAGGCACATTATATCAAATTCATTGT
TGAAGATATGAGTTATGAGGTCACAAATGTTGTCTTGTTACCCCGTGTGAGTAGTGACGTC
ATTTTCATGACTGAAATCTCTTCAACGCCGTTTAGCAATAATAGGCTCAGTAGTATTCAAC
CAATTACAATCAGTAGAAAATTCTCTATACTATTCTTATGTTGCATCCTGATATCCCTAT
GCAAAAATTAGTCATCTAATATAATCATTTTTCGATAAATACTTTGGGCAACAAATCAAT
GTAACATCTATTTTCTTTTCAG

DOMAIN 1C

CTACCTTTGAGGATGAAAAGCACAGCTTACGAATCAGAAAAAATGTCGACAGCTTGACTC
CTGAAGAAACAAATGAACTGCGTAAAGCCCTGGAGCTTCTTGAAAATGATCATACTGCAG
GTGGATTCAATCAGCTTGGCGCCTTCCATGGAGAGCCTAAATGGTGCCCTAATCCTGAAG
CGGAGACAAGGTTGCATGCTGTGTTTCATGGCATGGCTGTTTTCCCTCATTGGCACAGGC
TTCTTGCTCTCCAGGCGGAGAATGCTCTTAGAAAGCATGGGTACAGTGGTGCTCTACCAT
ACTGGGATTGGACTCGCCCCCTTTCCCAACTTCCTGATCTGGTTAGTCATGAGCAGTATA
CAGATCCTTCCGACCATCACGTGAAGCATAACCCGTGGTTCAATGGCCACATCGATACAG
TAAATCAGGATACCACCAGAAGCGTACGGGAGGATCTTTATCAACAACCTGAATTTGGAC

ATTTACGGATATTGCTCAACAAGTCCTCTTAGCATTAGAACAAGATGACTTCTGTTTCGT
TTGAAGTGCAGTATGAGATTTCCCATATTTTATCCATGCACTTGTAGGAGGAACCGACG
CTTATGGCATGGCATCGCTGAGATATACAGCATACGATCCAATCTTTTTCTTGCATCATT
CAAACACCGACAGGATCTGGGCTATTTGGCAATCCCTGCAAAAATACAGAGGCAAACCGT
ACAACACTGCCAACTGCGCCATAGAATCTATGAGAAGGCCCTGCAACCATTGTTGGACTAA
GCAGTGCCATTAAACCCTGACAGAATCACCAGAGAGCATGCTATCCCGTTTGATGTCTTCA
ACTATAGAGATAACCTTCATTACGTATATGATACCCTGGAATTTAATGGTTTGTTCGATTT
CACAACTTGATAGAGAGCTGGAAAAATCAAGAGTCACGAAAGAGTATTTGCTGGATTCT
TGCTGTCGGGGATTAAAAAATCTGCTCTTGTGAAATTCGAAGTTTGTACTCCACCTGATA
ATTGTCATAAAGCAGGGGAGTTTTATCTACTCGGGGACGAAAACGAGATGGCTTGGGCCT
ATGACCGACTTTTCAAGTATGATATTACTCAGGTTCTGGAAGCAAACCATCTACACTTCT
ATGATCATCTCTTCATTGCTACGAAGTCTTTGATCTTAAAGGAGTGAGTTTGGGAAGTG
ACCTGTTCCACACTGCAAATGTGGTACATGATTCCGGCACAG

INTRON 1C/1D (SEQ ID NO:116)

GTACGTGGATTTGATTACATAGCAATGCTATATGATTTTCAGTAATTACAACCTCAAGTCA
TGTAGCCGTTTTAGATTGCATTACATCAAACAGCATTGGATTAAATTGGGGGATTGTCCA
GGCCGCATTATGTTGCATTCCGAAAATAGTTTTGTGTCCAGTGTCCACGTTTAAAATTAAA
CCATTTTAATCATATTAGGGATAATTTAATAGATGTTATAGTGCTTTATTTTCATATTGT
TACAGTGGACAGTCACCAAGGACATATTTTACTCTATAGATACACAAACACCAATTAAAA
CCCTGCTTTGGAAAGTCTAACTTTTTCCCCACAG

DOMAIN 1D

GCACCCGTGATCGTGATAACTACGTTGAAGAAGTTACTGGGGCCAGTCATATCAGGAAGA
ATTTGAACGACCTCAATACCGGAGAAATGGAAAGCCTTAGAGCTGCTTTCCTGCATATTC
AGGACGACGGAACATATGAATCTATTGCCCAGTACCATGGCAAACCAGGCAAATGTCAAT
TGAATGATCATAATATTGCGTGTGTGTCCATGGTATGCCTACCTTCCCCCAGTGGCACA
GACTGTATGTGGTTCAGGTGGAGAATGCTCTCCTAAACAGGGGATCTGGTGTGGCTGTTT
CTTACTGGGAGTGGACTGCTCCCATAGACCATCTACCTCATTTTCATTGATGATGCAACAT
ACTTCAATTCCCGACAACAGCGGTACGACCCTAACCCCTTTCCTTCAGGGGAAAGGTTACTT
TTGAAAACGCAGTCACAACAAGGGACCCACAAGCCGGGCTCTTCAACTCAGATTATATGT
ATGAGAATGTTTTACTTGCCTGGAGCAGGAAATTTATTGTGACTTTGAAATTCAGTTTG
AGCTTGTTTCATAACGCACCTTCATTCCATGCTGGGAGGTAAAGGCGAGTACTCCATGTCCT
CCCTGGACTATTCTGCGTTTGATCCCGTCTTCTTCTTACATCATGCCAACACGGACAGAC
TGTGGGCAATCTGGCAGGAACACAAAGATTCCGAGAACTGCCTTATGAAGAAGCGAACT
GTGCAATCAACCTCATGCATCAACCACTGAAGCCGTTTCAGTGATCCACATGAGAATCACG
ACAATGTCACTTTGAAATACTCAAAACCACAGGACGGATTTCGACTACCAGAACCACTTCG
GATACAAGTATGACAACCTTGAGTTCATCACTTATCTATCCCAAGTCTTGATGCTACCC
TGAAGCAAAGGAGAAATCACGACAGAGTGTTTGCGGGCTTCCTTCTTCATAACATAGGAA
CTTCTGCTGACATAACTATCTACATATGTCTGCCTGACGGACGGCGTGGCAATGACTGCA
GTCATGAGGCGGGAACATTCTATATCCTCGGAGGCGAAACAGAGATGCCTTTTATCTTTG
ACCGTTTGTATAAATTTGAAATCACCAAACCACTGCAACAGTTAGGAGTCAAGCTGCATG
GTGGAGTTTTTCGAACTGGAGCTTGAGATCAAGGCATACAACGGTTCCTATCTGGATCCCC
ATACCTTTGATCCAACCTATCATCTTTGAACCTGGAACAG

INTRON 1D/1E (SEQ ID NO:117)

GTAATGCCATCTTAATACAGTTCGTTTCGTAAATTATATATGTTTCGTTTACAACACCATA
CCTTGAATTGAGGTAATACATCACTTGATATTGATAATGTAATGGTAATTGTTCTTGTTT
GTAAAACCGTTTTCTGGGGTGTATTACTATCCACCTGGTGGATAGTGAGTAAACACAT
TCGGTTTAAATATGGGTATCTAATGGACAGTGAAGTGTGCTGGCTAGGCAGATACCTTGGT
TTCTGTGAATGGAGGTAGTAGAAAGGGGTTTTGATGATTGCAG

DOMAIN 1E

ATACCCATATCTTGGACCACGACCATGAGGAAGAGATACTTGTCAGGAAGAATATAATTG
ATTTGAGCCCAAGGGAGAGGGTTTCTCTAGTCAAAGCTTTGCAAAGAATGAAGAATGATC
GCTCCGCTGATGGGTACCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGTCCCA
ATCCATCTGCAGCTCACCGTTATGCTTGCTGTGTCCATGGCATGGCTACATTTCCCCAGT
GGCACAGACTGTACACTGTTTCAAGTTTCAAGGATGCCCTGAGGAGACATGGTTCACTTGTTG
GTATTCCTTACTGGGACTGGACAAAACCAGTCAACGAGTTACCCGAGCTTCTTTCTTCAG
CAACATTTTATCATCCAATCCGGAATATTAATATTTCAAATCCATTCTTCGGGGCTGACA
TAGAATTTGAAGGACCGGGCGTTCATACAGAGAGGCACATAAATACTGAGCGCCTGTTTC
ACAGTGGGGATCATGACGGATACCACAACTGGTTCTTCGAAACTGTTCTCTTTGCTTTGG
AACAGGAAGATTACTGCGATTTTGAAATACAATTTGAGATAGCCCATAAATGGCATCCACA
CATGGATTGGTGGAAGCGCAGTATATGGCATGGGACACCTTCACTATGCATCATATGATC
CAATTTTCTACATCCACCATTACAGACGGACAGAATATGGGCTATTTGGCAAGAGCTGC
AGAAGTACAGGGTCTATCTGGTTCGGAAGCAAACGTGTCATTGAACATATGAGAACAC
CCTTGAAGCCTTTCAGCTTTGGGCCACCTTACAATTTGAATAGTCATACGCAAGAATATT
CAAAGCCTGAGGACACGTTTGACTATAAGAAGTTTGGATACAGATATGATAGTCTGGAAT
TGGAGGGGCGATCAATTTCTCGCATTGATGAACCTTATCCAGCAGAGACAGGAGAAAGACA
GAACTTTTGCAGGGTTTCTCCTTAAAGGTTTGTGATCATCCGCATCTGTGTCAATTGCAAG
TTTGCAGAGTTGATCACACCTGTAAAGATGCGGGCTATTTCACTATTCTGGGAGGATCAG
CCGAAATGCCATGGGCATTCGACAGGCTTTATAAGTATGACATTACTAAACTCTTCACG
ACATGAACCTGAGGCACGAGGACACTTTCTCTATAGACGTAACCTATCACGTCTTACAATG
GAACAGTACTCTCGGGAGACCTCATTACAGACGCCCTCCATTATATTTGTACCTGGACGCC

INTRON 1E/1F-1 (SEQ ID NO:118)

GTGAGTACCTGTTTGCACCTAAGACTTCTGTAGGCTAAAAGTGTAAGAAATATCAATTAAT
TTCAATTCACCCAAACTTGAAAACGGTACCTATATAGGTTAACTTTTTGTCTACAGTAA
CTGAACATACCTACACATTTTCATGAAATGATCTCTCAATATTTTCCACCAACAG

DOMAIN 1F-1 (1st part of domain f)

ATAAACTCAACTCACGGAAACATACACCTAACAGAGTCCGCCATGAGCTAAGTAGCCTTA
GTTCCCGTGACATAGCAAGCTTGAAGGCAGCTTTGACAAGCCTTCAACATGATAATGGGA
CTGATGGTTATCAAGCTATTGCTGCCTTCCATGGCGTTCTGCGCAGTGCCACGAGCCAT
CTGGACGTGAG

INTRON 1F-1/1F-2 (SEQ ID NO:119)

GTAAATTTACAGAGCTTTATGAAGTGTGTTTCAAGTGAAGAGACCAAGATATACTTATAC
CCAAAACCTAGCTAGCAACAGACGATTTCACTTGTTTCGGACACTTTGTATTATACGTTGG
ATCCCAAGGTAAACGGAAACGTAACCGAGAATCAGTCCGTAAAGTGAGTGAGTGAGTTTG
GGGCTTAACGTCGCACTCAGCAATACCCAGCTATGTGGCGACTCTCAGATTTACTGCTG
GAGGAGAACCTACATAGCCCGGTTTAAACCGTGTGGTATGTAGTAAGACCAGCGCGGCAT
GGCTGGTATCTGACGGACGAAGGGTGGCGCTGCACGTATTCAGTGGTACAACACTGCAC
CCCAATTTACCGACCGGAGAACTGATCTCCCCTTCGGAGATATCGCCTGCCTTCCACGG
GATTGCAACTCGGTGACCTTCAAGCCAGCGCGCTTCTAGCGGGGGCGATTAGAGGTTNAA
GGCCGACGGCTCTACCACCTTAACTATCCCCCGGCCCACTCCTGACGGAAATGTTTATA
ATTGAGCCTTTGTTTTCTTATTAAACACTCTTGGCAGATTTTCTATAGATAATGGATTCA
CATGTAGACAGTCTCCCATTTGTTGTAACCTGGTAGTCAAGAGTTAGAATCTGAATACATTC
TCCAAGATGGATCAAGGAAAACAATAATTACTTGATGTTGCAG

15/44

DOMAIN 1F-2 (2nd part of domain f)

ATCGCCTGTTGCATCCACGGCATGGCGACGTTTCCTCACTGGCACCGGTTGTACACTCTG
CAGTTGGAGCAAGCGCTGCGCAGACACGGGTCCAGTGTGCTGTTCCATACTGGGACTGG
ACCAAGCCAATCACCGAACTGCCACACATTCTGACAGACGGAGAATATTATGACGTTTGG
CAAAATGCCGTCTTGGCCAATCCGTTTGAAGAGGTTATGTGAAAATTAAAGATGCATTT
ACGGTGAGAAATGTCCAGGAAAGTCTGTTCAAAATGTCAAGTTTTGGAAAGCACTCGCTT
CTGTTTGACCAGGCTTTGTTGGCTCTTGAACAACTGACTACTGTGACTTCGAAGTTCAG
TTTGAAGTGATGCATAACACGATCCATTATCTCGTAGGAGGGCGTCAAACGTACGCCTTC
TCCTCTCTCGAGTATTCTCATACGATCCAATCTTCTTTATTACCACTCGTTTGTGAC
AAAATATGGGCTGTATGGCAAGAACTGCAAAGCAGGAGACATCTACAGTTTAGAACAGCT
GATTGTGCTGTGGGCCTCATGGGTCAAGCAATGAGGCCTTTCAACAAGGATTTCAACCAC
AACTCGTTCACCAAGAAGCACGCAGTCCCTAATACAGTATTTGATTATGAAGATCTTGGC
TATAACTATGACAACCTTGAAATCAGTGGTTTAAACTTAAATGAGATCGAGGCGTTAATA
GCAAACGCAAGTCACATGCTAGAGTCTTTGCTGGGTTCTGTTGTTTGGATTAGGAACT
TCGGCTGATATACATCTGGAAATTTGCAAGACATCGGAAAACCTGCCATGATGCTGGTGTG
ATTTTCATCCTTGGAGGTTCTGCAGAGATGCATTGGGCATACAACCGCCTCTACAAGTAT
GACATTACAGAAGCATTGCAGGAATTTGACATCAACCCTGAAGATGTTTTCCATGCTGAT
GAACCATTTTTCTGAGGCTGTCGGTTGTTGCTGTGAATGGAAGTGTATTCCATCGTCT
CATCTTCACCAGCCAACGATAATCTATGAACCAGGCGAAG

INTRON 1F-2/1G-1 (SEQ ID NO:120)

GTGAGATATATGCAAATTGAATGTTGTCCAGATGCGTTGTTTACATTTATATGCTTGGAA
TTGTCCTGAACGAATACAGTGGATAACCAAAAAGCTGAAAAATAAAAAGATATATACTTC
ATTCTGAATTTGTGAGTATTGCTGACCCAAAACACGTTATCCATGTGACACTATATTT
GCCTTTCTGAATCTGAGACTGCGTTATGTTTCTAATAATCACGAAATATGGTATACAGGT
TGTGTATCTGTAGAATAACCAAGGCAGAATTTAAAGGGTCACACCCTGTTTAATACAG

DOMAIN 1G-1 (1st part of domain g)

ATCACCATGACGACCATCAGTCGGGAAGCATAGCAGGATCCGGGGTCCGCAAGGACGTGA
ACACCTTGACTAAGGCTGAGACCGACAACCTGAGGGAGGCGCTGTGGGGTGTATGGCAG
ACCACGGTCCCAATGGCTTTCAAGCTATTGCTGCTTTCCATGGAAAACAGCTTTGTGTC
CCATGCCTGATGGCCACAACCTACTCATGTTGTACTCACG

INTRON 1G-1/1G-2 (SEQ ID NO:121)

GTAAGTTTGTGTTGGTTAGTGTGTTGTCATGTTTTGCCATATCGATAGTATCAGTGTGG
TAACATCTGGTTTCTAGTTCATTGCTTACCTTATCAGAAGCTGTTTGTCTCTCGTCTAC
AATAGTGACGTCTTTTCAAGTTTGAACCGTGTACATCCGGGTTATATTGGTCTCCAGCAA
CCCGTGCTTGTGCTGGGAGGCCACTGATGGGAACGGGTGGTCAGACTCGCTCACTTAGTT
GACACATGTCAATTGCGAAGATCGATGCTGAGGTTGTTAAACATTGGATTGTCTGGTCCA
GACTCGATTATTTACAGACAGCCGCCATGTACCTGGAATATTGCTGAGTGGGCGTTAAA
CAACAACTAGTCAGACTAATCTTTCACTGTTTATAATGATGGCTCGAACCTAGCACTCA
TGTCCCAAGTTGGCGAACATCTGGAAGGGAATTTCAAATGAAAAGAACAATCTTTCACGT
CTATTGGTATCACGCTCCTGGAGAAGAACATGATGTTACGGCGTTACTTCTCTTACCT
GTTTTACTTGTTCACGTTTCTTTCATATTTAAAGAGTATTTGGGTATTAGAGCTTTGGT
GCTGTTACAATGCTACTCAACTGTTTCAAGTGCAGGCGACCGCGCTTGTTCACATTAAGT
TTTGTGTTGTTGGTTGGTTTGTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTA
TGTGTGTGTGTGTGTATCTATGTCTATGTGTCTGTGTCTGTGTGTCTGTCTATGTGTGTG
TGTGTCTGTGTCTATGTGTGTGTCTGCGTGTGTGTCTGTGTCCGTATGTGGCTGTGTCTA
TGTGTGTGTGTGTCTGTGTTTATGTGTGTATATGCGTGTGTGTCTGTGTCCGTATGTGGC

TGTGTCTATGTGTGTGACATGCAATACATGCTGTGATACTCACTAGCTGCGTCTATCGAC
CAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACCTTCCCACACTGGCATCGCCTCTACACCAAGCAGATGGAGGATGCAATGA
GGGCGCATGGGTCTCATGTTCGGCCTGCCCTACTGGGACTGGACTGCTGCCTTCACCCACC
TGCCAACACTGGTCACCGACACGGACAACAACCCCTTCCAACAT

INTRON 1G-2/1G-3 (SEQ ID NO:122)

GTAAGAGCGGGGTAGGGATGGGGTGGTAGGGGGTGGGTGTCTATTACTTCCCGCTTCA
CTTGTATGAAATGGATAACCTTGGCTGCATCCCAATTGCGTGATCGATTCTCTTTCGATT
CACTCGTGCGATTAGACTGCCTTATTTACTATAGTAGTTAGAATGTTGCTCAGTGCGCCG
TTAAACAACATAACACAAAACCGCATTGTGTTTTATATGGTCACTCTACTGTTTATCACG
TATATGTATGTTCCGACTCACTGGTTGGTGCGTACCATTCTACTGTCACACTGAGAGCCA
ATGTTCTCAGATGTGTGAAATGTTTGAAAGCCGTTTCTACATAATATTGCAGGAATACCA
TTGTAGAATGTAGTCAAACAGGTAACAATCTGTTAGTGAGCCAGTTCGAGGTTGCGTTG
TAGGGTGTAGTCCAACAGGTAGGCAGTCCATAAGCATAGTTTTTAAGCATTTTAGATCAT
CTATAATTAACCACATGGTTAGCCGCTATGTTTAGTTTAATCCAGTATAAGTTAGAAGTG
TTATATTTTCGAAGGGAAGTGAGTAAATCCTTATTCCTTGACTACCATTTAATAGATTTCC
CAATGACTCCATTCAACTCCTAACTTTTACATCACTGCTCTCTTCAACAG

DOMAIN 1G-3 (3rd part of domain g)

GGACACATTGATTATCTCAATGTCAGCACAACTCGATCTCCCCGAGACATGCTGTTCAAC
GACCCCGAGCATGGATCAGAGTCGTTCTTCTACAGACAAGTCCTCTTAGCTCTGGAACAA
ACTGATTTCTGCAAATTCGAAGTTCAGTTTGAGATAACCCACAATGCCATCCATTCCCTGG
ACAGGTGGCCACAGCCCCCTACGGAATGTCCACTCTCGACTTCACTGCCTACGATCCTCTC
TTCTGGCTTCACCACTCCAACACCGACAGAATCTGGGCTGTCTGGCAAGCTTTGCAAGAA
TACAGAGGACTTCCATACAACCATGCCAATTGTGAGATCCAGGCAATGAAAACGCCCCCTG
AGGCCTTTTCAGTGACGATATCAACCACAACCCAGTCACAAAGGCTAACGCGAAGCCATTA
GATGTGTTTCGAGTATAATCGGTTGAGCTTCCAGTACGACAACCTCATCTTCCATGGATAC
AGTATTCCGGAACCTTGATCGCGTGCTTGAAGAAAGAAAGGAGGAGGACAGAATATTTGCT
GCCTTCCCTTCTCAGTGGAATCAAGCGTAGTGCTGATGTAGTGTTTCGACATATGCCAGCCA
GAACACGAATGTGTGTTTCGCAGGGACTTTTGCGATTTTGGGAGGGGAGCTAGAAATGCCC
TGGTCCTTCGACAGACTGTTCCGCTATGATATCACCAAGGTGATGAAGCAGCTACACCTG
AGGCATGACTCTGACTTTACCTTCAGGGTGAAGATTGTTCGGCACCGACGACCACGAGCTT
CCTTCAGACAGTGTCAAAGCACCAACTATTGAATTTGAACCGGGCG

INTRON 1G-3/1H (SEQ ID NO:123)

GTGAGTACGACAGGCATTTCTAGTAAAAACCTACTTTTGGTAAAAGGTTTCGAGAAATCAC
TTGAAGCAACAACATGATTTTGTAAACGCCTATTACACGTGAACATGTCACACCCGGTGAT
GCCGTTTAATGGACATGCCTCTGTTAATGAAAGGGGTAAGTACATGTGTATGGGGATGGG
ATGGGAGCCACCTGTCCCAATTTTCATAGGTCCTTAGGATCCAGTTGCGTAGGAATCCCC
TGATTAATGCCTTGTGAATTCCTCCTGGAATTGTCCTGGCCCAAATTTTTACAAACCCGC
CCCGATATACCTTGGAATAATTGGGCCTAAGGGTGGGGCTTTTAAGGACCAAGAACCCA
ACCTAAACCCCAACCCATTTTTTCCCACCCATTCCAGGTTTTGTTTTACCAAATAAAAAG
GTTTCCACTTTGAGGAAACCCCTTTAAGGGTCTTTTTCAGGGCTTTTTTCTTTTCTGGGA
ATTCCAATTCCGGGGGAACAAAATACATATATTTACAGACCTTTGGTCAAATTTATATA
ATTTCCGACTTCATGTCATAGGTTTGTCTTTCTTCCTACACAG

DOMAIN 1H

TGCACAGAGGCGGAAACCACGAAGATGAACACCATGATGACAGACTCGCAGATGTCCTGA
TCAGGAAAGAAGTTGACTTCCTCTCCCTGCAAGAGGCCAACGCAATTAAGGATGCACTGT
ACAAGCTCCAGAATGACGACAGTAAAGGGGGCTTTGAGGCCATAGCTGGCTATCACGGGT
ATCCTAATATGTGTCCAGAAAGAGGTACCGACAAGTATCCCTGCTGTGTCCACGGAATGC
CCGTGTTCCCCCACTGGCACCGCCTGCATACCATTTCAGATGGAGAGAGCTCTGAAAAACC
ATGGCTCTCCAATGGGCATTCCCTTACTGGGATTGGACAAAGAAGATGTCGAGTCTTCCAT
CTTTCTTTGGAGATTCCAGCAACAACAACCCTTTCTACAAATATTACATCCGGGGCGTGC
AGCACGAAACAACCAGGGACATTAATCAGAGACTCTTTAATCAAACCAAGTTTGGTGAAT
TTGATTACCTATATTACCTAACTCTGCAAGTCTTGGAGGAAACTCGTACTGTGACTTTG
AAGTTCAGTATGAGATCCTCCATAACGCCGTCCACTCCTGGCTTGGAGGAACTGGAAAGT
ATTCCATGTCTACCCTGGAGCATTGGGCCTTTGACCCTGTCTTCATGATTACCACTCGA
GTTTGGATAGAATCTGGATCCTTTGGCAGAAGTTGCAAAAGATAAGAATGAAGCCTTACT
ACGCATTGGATTGTGCTGGCGACAGACTTATGAAAGACCCCCCTGCATCCCTTCAACTACG
AAACCGTTAATGAAGATGAATTCACCCGCATCAACTCTTTCCCAAGCATACTGTTTGACC
ACTACAGGTTCAACTATGAATACGATAACATGAGAATCAGGGGTCAGGACATACATGAAC
TTGAAGAGGTAATTCAGGAATTAAGAAACAAAGATCGCATATTTGCTGGTTTTGTTTTGT
CGGGCTTACGGATATCAGCTACAGTGAAAGTATTCATTTCATTCGAAAAACGATACAAGTC
ACGAAGAATATGCAGGAGAATTTGCAGTTTTGGGAGGTGAGAAGGAGATGCCGTGGGCAT
ATGAAAGAATGCTGAAATTGGACATCTCCGATGCTGTACACAAGCTTCACGTGAAAGATG
AAGACATCCGTTTTAGAGTGGTTGTTACTGCCTACAACGGTGACGTTGTTACCACCAGGC
TGTCTCAGCCATTCATCGTCCACCGTCCAGCCCATGTGGCTCACGACATCTTGGTAATCC
CAGTAGGTGCGGGCCATGACCTTCCGCCTAAAGTCGTAGTAAAGAGCGGCACCAAAGTCG
AGTTTACACCAATAGATTCGTCGGTGAACAAAGCAATGGTGGAGCTGGGCAGCTATACTG
CTATGGCTAAATGCATCGTTCCCCCTTTCTCTTACCACGGCTTTGAACTGGACAAAGTCT
ACAGCGTCGATCACGGAGACTACTACATTGCTGCAGGTACCCACGCGTTGTGTGAGCAGA
ACCTCAGGCTCCACATCCACGTGGAACACGAGTAG

3' UTR

TTCACAG

INTRON 3' UTR (SEQ ID NO:124)

GTGAGGAGAAGGCCCCAGGCTAGCAGGGCAATGGATGAAGGAAATAGGGGCAAAGGGAAT
AGCAGTTACACCATCGACATTTCCAACCTCCTCAGAACTAATATATAGCCTTAATACAA
CCAGCCAAGACTCAACGGGCAGCCGGGGTGGGGGGATTGTTGGTGGTCGCTGTTTCAGACCA
GGGTGCAAAATATCAGTGCGCAAATCAACATGTTGCGTGTGAGACACTGACACAGCAGTC
ATTGAACCTGCAGACCCATAACAGGAAATGGGGCAGATACGATCAAAGACAGTGTAATA
TAGGGATAAGTAGGCATATGCAACCACCTGATGGAAATGAAAAGGGGTAAGTTTAAACCC
CGGCTACCAAAGGTCCAATGGTTCCCTTAACCCAGCTTACGCTATCCCTCTAATTTAGTA
TTGAGCTGATTTCTGTGAGTTCATGTAAACTGTATACTTTCTGTATTATTACAG

3' UTR

GTTGCTATGCCGACTGCGCTATATTGGTGAACGAGACGATGAGGACATCTCTGAAAGAGT
TCGCCAAGTGATGTGTAGGTCACGGAAGTATTGTTGAGCTAACAATATGATGATTTCAA
ATGACTTGGCGCTCTAGGACAAAGACATAATTCATCAGCACCCGTGTGCACCAACTCTTTG
TTTGCTGCAACGTCTGACAAGCGACACGTCAATCAACAAGCTGTTCAAACCTCAAGTGA
TGTAAC TAGAATCGTTGGGCCATCGTTCACAAAGTATTGACAGATGTCACACATGATGGC
GAGAAACACTTTAGAACTTTTAATGACCTAGAGTGACTTGTAATATGTAAATATATTCT

TCAAAGACTCAGCTGAACTATTGTTGGATAACACATCAATTCCTCAACAAAATGCTTTA
TCTTCACATGGATGTATGTAATGTGGCCGGCAATAAAGTATATATATGTAT

Figure 5

Primary structure of the HtH1 protein

SIGNAL PEPTIDE

LVQFLLVALVVGAGA

DOMAIN A

DNVVRKDVSHLTVDEVQALHGALHDVTASTGPLSFEDITSYHAAPASCDYKGRKIACCVHGMPSFP
FWHRAYVVQAERALLSKRKTVMGPYWDWTQTLTHLPSLVTEPIYIDSKGGKAQTNYWYRGEIAFIN
KKTARAVDDRLFKEKVEPGHYTHLMETVLDALEQDEFCKFEIQFELAHNAIHVYLVGGKFEYSMSNLE
YTSYDPIFFLHHSNVDRLFIAIWQRLQELRGKNPNAMDCAHQQLQPFNRDSNPVQLTKDHSTP
ADLFDYKQLGYSYDSLNLNGMTPEQLKTELDERHRSKERAFAFSLSGFGGSANVVVYACVPDDDP
SDDYCEKAGDFFILGGQSEMPWRFYRPFYDVTEAVHHLGVPLSGHYVVKTELSVNGTALSPDLL
PQPTVAYRPGK

DOMAIN B

GHLDPVPHRRHDDDLIVRKNIDHLTREEEYELRMALERFQADTSVDGYQATVEYHGLPARCPRPDA
KVRFACCMHGMAFPHWHRLFVTQVEDALVRRGSPIGVPYWDWTKPMTHLPDLASNETYVDPYGHT
HHNPFFNANISFEEGHHHTSRMIDSKLFAPVAFGEHSHLFDGILYAFEQEDFCDFEIQFELVHNSI
HAWIGGSEDSMATLHYTAFDPIFYLHHSNVDRLWAIWQALQIRRHKPYQAHCAQSVEQLPMKPFA
FPSPLNNEKTHSHSVPTDIYDYEEVLHYSYDDLTFGGMNLEEIEEAHLRQQHERVFAGFLLAGI
GTSALVDIFINKPGNQPLKAGDIAILGGAKEMPWAFDRLYKVEITDSLKTLSLDVGDDYEVTFKIH
DMHGNAIDTDLIPHAHVSEPAH

DOMAIN C

PTFEDEKHSRLIRKNVDSLTPETNELRKALELLENDHTAGGFNQLGAFHGEPKWCPNPEAEHKVA
CCVHGMVFPWHRLALQAENALRKHGYS GALPYWDWTRPLSQLPDLVSHEQYTDPSDHHVKHNP
WFNGHIDTVNQDTRSVREDLYQQPEFGHFTDIAQQVLLALEQDDFCSFEVQYEISHNFIHALVGG
TDAYGMASLRYTAYDPIFFLHHSNTDRIWAIWQSLQYRGKPYNTANCAIESMRPLQPFGLSSAI
NPDRITREHAIPFDVFNRYRDNLHYVYDTLEFNGLSISQLDRELEKIKSHERVFAGFLLSGIKKSAL
VKFEVCTPPDNCHKAGEFYLLGDENEMAWAYDRLFKYDITQVLEANHLHFYDHLFIRYEVFDLKG
SLGTDLFHTANVVHDSGT

DOMAIN D

GTRDRDNYVEEVTGASHIRKNLNDLNTGEMESLRAAFLHIQDDGTYESIAQYHGKPGKCQLNDHNI
ACCVHGMPTFPQWHRLYVVQVENALLNRGSGVAVPYWEWTAPIDHLPHFIDDATYFNSRQQRYDPN
PFFRGKVTFENAVTTRDPQAGLFNSDYMYENVLLALEQENYCDFEIQFELVHNALHSM LGKGQYS
MSSLDYSAFDPVFFLHHANTDRLWAIWQELQRFREL PYEEANCAINLMHQPLKPFSDPHENHDNVT
LKYSKPQDGFYQNHFGYKYNLEFHLSIPSLDATLKQRRNHDRVFAGFLLHNIGTSADITIYIC
LPDGRRGNDCSHEAGTFYILGGETEMPFI FDRLYKFEITKPLQQLGVKLHGGVFELELEIKAYNGS
YLDPHTFDPTIIFEPGT

DOMAIN E

DTHILDHDHEEEILVRKNIIDLSPRERVS LVKALQRMKNDRSADGYQAIASFHALPPLCPNPSAAH
RYACCVHGMATFPQWHRLYTVQVDALRRHGS LGVGPYWDWTKPVNELPELLSSATFYHPIRNINI
SNPFLGADIEFEGPGVHTERHINTERLFHSGDHDGYHNWFFETVLFALQEDYCDFEIQFEIAHNG

IHTWIGGSVAVYGMGHLHYASYDPIFYIHHSQTDRIWAIWQELQKYRGLSGSEANCAIEHMRTPLKP
FSFGPPYNLNSHTQEYSKPEDTFDYKKFGYRYDSLELEGRSISRIDELIQQRQEKDRTFAGFLLKG
FGTSASVSLQVCRVDHTCKDAGYFTILGSSAEMPWAFDRLYKYDITKTLHDMNLRHEDTFSIDVTI
TSYNGTVLSGDLIQTPSIIFVPGR

DOMAIN F

HKLNSRKHTPNRVRHELSSSLSSRDIAASLKAALTSLQHDNGTDGYQAIAAFHGVPAQCHEPSGREIA
CCIHGMATFPHWHRLYTLQLEQALRRHGSSVAVPYWDWTKPITELPHILTDGEYYDVWQNAVLANP
FARGYVKIKDAFTVRNVQESLFBKMSSFGKHSLLFDQALLALEQTDYCDFEVQFEVMHNTIHYLVGG
RQTYAFSSLEYSSYDPIFFIHHSFVDKIWAVWQELQSRRHLQFRTADCAVGLMGQAMRPFNKNDFNH
NSFTKKHAVPNTVFDYEDLGYNYNLEISGLNLNEIEALIAKRKSHARVFAGFLLFGLGTSADIHL
EICKTSENCHDAGVIFILGSSAEMHWAYNRLYKYDITEALQEFDINPEDVFHAEPPFLRLSVVAV
NGTVIPSSHLHQPTIIYEPGE

DOMAIN G

DHHDDHQSGSIIAGSGVRKDVNTLTKAETDNLREALWGVMAHGPNGFQAIAAFHGKPALCPMPDGH
NYSCTHGMATFPHWHRLYTKQMEDAMRAHGSHVGLPYWDWTAAFTHLPTLVTDTDNNPFQHGHI
YLVNSTTRSPRDMLENDPEHGSESFFYRQVLLALEQTDFFCKFEVQFEITHNAIHSWTGGHSPYGM
TLDFATYDPLFWLHHSNTDRIWAVWQALQFYRGLPYNHANCEIQAMKTPLRPFSDDINHNPTKAN
AKPLDVFEYNRLSFQYDNLI FHGYSIPELDRVLEERKEEDRIFAAFLLSGIKRSADVDFDICQPEH
ECVFAGTFAILGGELEMPWSFDRLEFRYDITKVMKQLHLRHDSDFTRVKIVGTDDHELPSDSVKAP
TIEFEPG

DOMAIN H

VHRGGNHEDEHHDDRDLADVLIRKEVDFLSLQEANAIDALYKQLQNDDSKGGFEAIAGYHGYPNMCP
ERGTDKYPCCVHGMPVFPHWHRLHTIQMERALKNHGSPMGIPYWDWTKMSSLPSFFGDSSNNNPF
YKYYIRGVQHETTRDINQRLFNQTKFGEFDYLYLTLQVLEENSYCDFEVQYEILHNAVHSLWGGT
GKYSMSTLEHSAFDPVFMIIHSSLDRIWILWQKLQKIRMKPYALDCAGDRIMKDPLHPFNYETVN
EDEFTRINSFPSILFDHYRFNYEYDNMRIRGQDIHELEEVQELRNKDRIFAGFVLSGLRISATVK
VFIHKNDSHEEYAGEFAVLGGEKEMPWAYERMLKLDISDAVHKLHVKDEDIRFRVVVTAYNGDV
VTTRLSQPFIVHRPAHVAHDILVIPVGAGHDLPPKVVKSGTKVEFTPIDSSVNKAMVELGSYTAM
AKCIVPPFSYHGFELDKVYSVDHGDYIIAAGTHALCEQNLRLHIHVEHE

Figure 6

Genomic sequence of the HtH2 gene

DOMAIN 2A-1 (1st part of domain a)
[domain a, parts 1-4: SEQ ID NO:156]

GGTCTTCCGTACTGGGACTGGACGCAGCATCTGACTCAACTCCCAGATCTGGTGTCTAGACCCCTTG
TTTGTCGACCCGGAAGGAGGAAAG

INTRON 2A-1/2A-2 (SEQ ID NO:125)

[illegible]

DOMAIN 2A-2 (2nd part of domain a)

GCCCATGACAACGCATGGTATCGTGGAACATCAAGTTTGAGAATAAGAAGACTGCAAGAGCTGTT
 GACGATCGCCTTTTCGAGAAGGTTGGACCAGGAGAGAATACCCGACTCTTTGAAGGAATTCTCGAT
 GCTCTTGAACAGGATGAATTCTGCAACTTCGAGATCCAGTTTGAGTTGGCTCACAAACGCTATCCAC
 TACCTGGTTGGCGGCCGTCACAC

INTRON 2A-2/2A-3 (SEQ ID NO:126)

GTGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATG
GTCACGAGTCACGTTCTCTGATGGTCACGAGTCACATTCTCTGATGGTCACGAGTCACATTCTCTG
TTGAGTGAAGTCTCAGTACCATTTATTTCTCTTACCTTCTTCTAACCAGGGGTTTCAGCGTGGATC
GTCTGAGAAGTTAGCGCAAATCTATATTGAAGTCATTTTTCTATCATATAACCATCGTTATATCCA
CGTGCGAAAGTGTTCAATTAATTATTTTTATTTTCATTTATGAAGGTCTAAAAGAAAATATGTATTG
TTGGAAACTATATTGCAAGGTGAAGGCAACACGAGTGTATTAATATTCTCAATATCAATGTACGCT
CTGTCAGCACCTGTTTCACCAGGAACACCTTTAGCGTACCAAATATCAGCTGATGATTTCGA
AGCGGACTATACCTCACCACCTTGTTTTGTGTGTGATTTATGTGTGCATGTGTGTGCGTGCGTGC
GTGTGTGTGTGTGTCCTACGTATGTTGATATTTTGTCTGACTGTATATGTTCTGCTTACCATTG
AAG

DOMAIN 2A-3 (3rd part of domain a)

GTACTCCATGTCTCATCTCGAGTACACCTCCTACGACCCCCCTCTTCTTCCCTCCATCACTCCAACAC
CGACCGCATCTTCGCCATCTGGCAACGTCTTCAGGTACTCAGAGGAAAGGACCCCAACACCGCCGA
CTGCGCACACAACCTCATCCATGAGCCCATGGAACCGTTCCGTCGGGACTCGAACCCTCTTGACCT
CACCAGGGAAACTCCAAACCAATTGACAGCTTTGATTATGCCACCTTGGCTACCA

INTRON 2A-3/2A-4 (SEQ ID NO:127)

GTATGTATGATTCTAATAATGAATGTTTTTACCTCCGGTTTAAACAATATTTTAGTATTACGAAAG
GAGAAGTACCTCGAGAGGTCTAGGTCTCAGATGTTTAGAAACCCATGAAGACAGGTATGCTTCTGA
AAAACAAAGTAACATCATGAGGCTAAAGTTCAGATTCAAACCATCGTAGTTTGAATCCAGCATGCA
AAGGGCCCTAACCCTGTAGATGGCGCTGCTTGAAACAGAGTAGTCTGTTTCAAGGGTCAGTACTGTCC
CCACAAACATCATAGTCAGGGTCAGTACTGTCCCCACAAACATCATAGTCAGGGTCAGTACTGTCC
CCACAAACATCACAGTCAGGGTTAATTTTGGATTTCGTTTTTGAATGCGAAGAAGACAGTCACGCCC
TGACACTGGACCGAGGTTGCCGAGAAAGCTCGTGATATTGCTGGAATACTGCCCAGTAAAACCATC
ATTTATTTTAGGCTATTTATTACGAAAAATAATAATATGTATAGAAATGCATATGATCGCTGTTTG
AATGTAAAATTTAGAATGGGTTTGGGAGTGTTCACTATTTTTTCATCAAATTTTCATGTATTTTAA
CCGATCGACGCTGAAGACAACTACCGTTAATCAGGCAGTTCATTCATATCTGATAGGGAATATTG
GTTGTTAACCAACGCTACATTGTGTCCAG

DOMAIN 2A-4 (4th part of domain a)

GTATGATGACTTGACCCTGAACGGTATGACCCCAGAGGAATTGAACTCATATCTGCATGAACGGTC
AGGCAAGGAGGGGGTGTTCGCAAGCTTCCGACTCTCAGGTTTTGGCGGCTCTGCTAACGTTGTTGT
CTACGCATGCCGTCTGCCACGATGAAATGGCTGTTCGATCAGTGCGACAAAGCCGGCGACTTCTT
TGTGTTGGGCGGACCCACCGAGATGCCCTGGAGGTTTTACAGAGCATTCCACTTCGACGTCACCGA
CAGCATCGACAACATCGACAAGGACCGCCACGGCCACTATTATGTAAAGGCGGAATTATTCAGTGT
AAATGGAAGTGCGCTACCGAATGATCTCCTGCCTCAACCCACCATCTCACACAGGCCAGCCCGCGG
ACACGTTGATG

INTRON 2A-4/2B (SEQ ID NO:128)

GTAAATGGCCATTGTATACATGCATTTCATTTGGACTTTGAGTGAGTGAGTGGATGCGTATTCAGTA
AGTGAGAGTGTGAGTGGGTATTAGGTCTGTGAGTGGGTTGGTGAGTGGATGGGTGAGTAAGAGTGG
GTTGGTGAGAAAGTGAGTGAGTCACTTGGTGGGTGCGTTAGTGGAAGCGTGATTGAGTGGATGGGA
GGTAGGTGAGTGAGTGAATTGGTGGGGGGGTGAGTGAGGTTAACGCTGTTCTGCTGTTCAATCACA
CCACATGTTGCCAGCTTACTGTGCAGGACGAATCCAGGGTTGTGTTAAATTTTATATGTTTATATA
TAACGATGGACGTGTCTGGATGTGGCGAATGTGTCAAGAGAATTATGCGGCTTTGTGCTGCTCCGC
GTATTTATTGCACGCGCGTTGGTACGCGGTTGATAAAGTAGTTCAAACATTTCCCAGCCATCTTT
GTCTGTTGTGAAAACCTACTCCAGGACCATCCATTTCAATATGTGTCTGCGTTTCATGGAGTTATAC
ATGTTAAACTGTAGAGCGCAGATGAGCACACTTGAGCATTTCTTCAGTAAATCAGAATGTGTATAT
TTCAAATTTACCAAATGCAATATCATCAAGCAAATTATGCAGCTCTATAGTAACATCGGAGTCAA
TGGTCCAGTGTGCCCTCGGCTGCCATTCCGACCTCCCTGGCCAGAATACACCCCGGTCAGGATCAG
TTATCCGTCAGAAGGCACGGTGCGGAATGAAAACATAAACACATAGTCGCTTAGTAGTATGCTGAT
TTAGGCACGCAAAATCCGAATGTGAATTACTGTGAATTGCATTACCTGTTACAG

DOMAIN 2B

AGGCCCCAGCTCCCTCCTCGGATGCTCACCTCGCCGTCAGGAAGGATATCAACCATCTGACACGCG
AGGAGGTGTACGAGCTGCGCAGAGCTATGGAGAGATTCCAGGCCGACACATCCGTTGATGGGTACC
AGGCTACGGTTGAGTATCACGGCTTACCTGCTCGATGTCCATTCCCCGAGGGCCACAAATAGGTTTCG
CCTGTTGCATCCACGGCATGGCGACATTCCTCATTGGCACAGACTGTTTCGTTACCCAGGTGGAAG
ATGCACTGATCAGGCGAGGATCCCCTATAGGGGTCCCCTACTGGGACTGGACTCAGCCTATGGCAC
ATCTCCCAGGACTTGCAGACAACGCCACCTATAGAGATCCCATCAGCGGAGACAGCAGACACAACC
CGTTCCACGATGTTGAAGTTGCCTTTGAAAATGGGCGTACAGAAGGTCACCCAGATAGTAGATTGT

TTGAACAACCTCTATTTGGCAAACATACGCGTCTCTTCGACAGTATAGTCTATGCTTTTGAGCAGG
AGGACTTCTGCGATTTTGAAGTTCAATTTGAGATGACCCATAATAATATTCACGCCTGGATTGGTG
GCGGCGGGAAGTATTCCATGTCTTCTCTACACTACACAGCCTTCGACCCTATCTCCTACCTTCATC
ACTCCAACACTGACCGTCTCTGGGCAATTTGGCAAGCGTTGCAGATACGAAGAAACAAACCGTATA
AGGCTCATTGTGCTTGGTCTGAGGAACGCCAGCCTCTCAAACCTTTCGCCTTCAGTTCCCCACTGA
ACAACAACGAAAAACCTACGAAAACCTCGGTGCCACCAACGTTTACGACTACGAAGGAGTCCTTG
GCTATACTTATGATGACCTCAACTTCGGGGGCATGGACCTGGGTGACGTTGAGGAATACATCCAGA
GGCAGAGACAGAGAGACAGGACCTTTGCTGGCTTCTTTCTGTACATATTGGTACATCAGCGAATG
TTGAAATCATTATAGACCATGGGACTCTTCATACCTCCGTGGGCACGTTTGCTGTTCTTGCGGAG
AGAAGGAGATGAAATGGGGATTTGACCGTTTGTACAAATATGAGATTACAGATGAACTGAGGCAAC
TTAATCTCCGTGCTGATGATGGTTTCAGCATCTCTGTTAAAGTAACTGATGTTGATGGCAGTGAGC
TGTCCTCTGAACTCATCCCATCTGCTGCTATCATCTTCGAACGAAGCCATA

INTRON 2B/2C (SEQ ID NO:129)

GTAAGTAGCTACCTGTTTATTCAATTTTTTCGCTTTGCCAATCAATTCATTTCAGCTTGAAATTCAA
TAATTGTGTTTTGCATGGCTGAAAACCAATTTGAACTCTTTTCTTTTCTCAGGTGCAACTCAAATA
AATAATCACTAATTGTTATGCACGCGGGTAGGGCATACTATATCCACATCGGTCACTCTCAA
ATGCAAACAAATTGTCTTATTTCCGTTGGGACAAGCAAACCCCTTTCCTGTAATCTTGCTTTGG
CATCCACTGGAATTAATGTTGACTGGTAATTGATACTGGCTCTCTTCTTGCATAGAGTTAATATCT
ATAGTTTGTAATCTTTATGATTTTGCTATTTATATTTTCGACAGCATGCTATAGACACCCTAGACT
ATTGTATAGCCACTGTATTGTTTTCCATTTATTATTTATAACAGAACATGGCTTGTAATTTTA
TTTACCTTCCAG

DOMAIN 2C

TTGACCATCAGGACCCTCATCAGGACACAATCATCAGGAAAAATGTTGATAATCTTACACCCGAGG
AAATTAATTCTCTGAGGAGGGCAATGGCAGACCTTCAATCAGACAAAACCGCCGGTGGATTCCAGC
AAATTGCTGCTTTTACGGGGAACCCAAATGGTGCCCAAGTCCCGATGCTGAGAAGAAGTTCTCCT
GCTGTGTCCATGGAATGGCTGTCTTCCCTCACTGGCACAGACTCCTGACCGTGCAAGGCGAGAATG
CCCTGAGAAAGCATGGATGTCTCGGAGCTCTCCCTACTGGGACTGGACTCGGCCCCGTGTCTCACC
TACCTGATTTGGTAAGTCAGCAGAACTACACCGATGCCATATCCACCGTGGAAGCCCGAAACCCCT
GGTACAGCGGCCATATTGATACAGTTGGTGTGACACAACAAGAAGCGTCCGTCAAGAACTGTATG
AAGCTCCCGGATTTGGTCAATTATACTGGGGTCGCTAAGCAAGTGCTTCTGGCTTTGGAGCAGGATG
ACTTCTGTGATTTTGAAGTCCAGTTTGAGATAGCTCACAATTTTCATCCACGCTCTTGTCGGCGGAA
GCGAGCCATATGGTATGGCGTCACTCCGTTACACTACTTATGATCCAATTTTCTACCTCCATCATT
CTAACACTGACAGACTCTGGGCTATATGGCAGGCTCTACAAAAGTACAGGGGCAAACCTTACAATT
CCGCCAACTGTGCCATTGCTTCTATGAGAAAACCCCTACAGCCCTTTGGTCTGACTGATGAGATCA
ACCCGGATGATGAGACAAGACAGCATGCTGTTCTTTTCAGTGTCTTTGATTACAAGAACAACCTCA
ATTATGAATATGACACCCCTTGACTTCAACGGACTATCAATCTCCAGCTGGACCGTGAAGTGTAC
GGAGAAAGTCTCATGACAGAGTATTTGCCGGATTTTGTGCTGCATGGTATTCAGCAGTCTGCACTAG
TTAAATTCTTTGTCTGCAATCAGATGATGACTGTGACCACTATGCTGGTGAATTCTACATCCTTG
GTGATGAAGCTGAAATGCCATGGGGCTATGATCGTCTTTACAAATATGAGATCACTGAGCAGCTCA
ATGCCCTGGATCTACACATCGGAGATAGATTCTTCATCAGATACGAAGCGTTTGATCTTCATGGTA
CAAGTCTTGGAAGCAACATCTTCCCCAACCTTCTGTCATACATGACGAAGGGGCAG

INTRON 2C/2D (SEQ ID NO:130)

GTGAGAACATTGATAATAGTTCAAATGAAGTATATCCGATTCAAGCTGTCGATACAAGATGAGATA
CATAATCACAATGTTTGTATTAGATATCTCTCTTAATTTAATGCCGCTTTTATCAATATTCGAGCA
ATCCTTCAGCAACATACACCAGCAAATGTTTCATCAACAGACTATATTATTTAATATTTTAAAT
CCTTCTCTGTTGTTATAAATACTTAAAGTATCGAATTCCTTGAATGCGTCTTCTCTGCAGCATATA
GTTAAGTTGTTGTGTTTCTCTGTCTG

DOMAIN 2D

GTCACCATCAGGCTGACGAGTACGACGAAGTTGTAAGTCTGCTGCAAGCCACATCAGAAAGAATTTAA
AAGATCTGTCAAAGGGAGAAGTAGAGAGCCTAAGGTCTGCCTTCTGCAACTTCAGAACGACGGAG
TCTATGAGAATATTGCCAAATTCCACGGCAAGCCTGGGTTGTGTGATGATAACGGTCGCAAGGTTG
CCTGTTGTGTCCATGGAATGCCACCTTCCCCCAGTGGCACAGACTCTATGTCCTCCAGGTGGAGA
ATGCTTTGCTGGAGAGAGGATCTGCCGTCTCTGTGCCATACTGGGACTGGACTGAAACATTTACAG
AGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCAATTCCCGTCAACAAACGTTTGACCCTAATC
CTTCTTTCAGAGGTAATAATCAGTTTTGAGAATGCTGTTACAACACGTGATCCCCAGCCTGAGCTGT
ACGTTAACAGGTACTACTACCAAAACGTCATGTTGGCTTTTGAACAGGACAACACTACTGCGACTTCG
AGATACAGTTTGAGATGGTTCACAATGTTCTCCATGCTTGGCTTGGTGGAAGAGCTACTTATTCTA
TTTCTTCTCTTGATTATTCTGCATTCGACCCTGTGTTTTTCTTTCACCATGCGAACACAGATAGAT
TGTGGGCCATCTGGCAGGAGCTGCAGAGGTACAGGAAGAAGCCATAACAATGAAGCGGATTGTGCCA
TTAACCTAATGCGCAAACCTCTACATCCCTTCGACAACAGTGATCTCAATCATGATCCTGTAACCT
TTAAATACTCAAAACCCACTGATGGCTTTGACTACCAGAACAACCTTTGGATACAAGTATGACAACC
TTGAGTTCAATCATTTCAGTATTCCCAGGCTTGAAGAAATCATTCGTATTAGACAACGTCAAGATC
GTGTGTTTGCAGGATTCCCTCCTTCACAACATTGGGACATCCGCAACTGTTGAGATATTCGTCTGTG
TCCCTACCACCAGCGGTGAGCAAAACTGTGAAAACAAAGCCGGAACATTTGCCGTACTCGGAGGAG
AAACAGAGATGGCGTTTTCATTTTGACAGACTCTACAGGTTTGACATCAGTGAAACACTGAGGGACC
TCGGCATACAGCTGGACAGCCATGACTTTGACCTCAGCATCAAGATTCAAGGAGTAAATGGATCCT
ACCTTGATCCACACATCCTGCCAGAGCCATCCTTGATTTTTGTGCCTGGTTCAA

INTRON 2D/2E (SEQ ID NO:131)

GTAAGAAAGTTTCACTGTCTAAATCTTTTTTTATGATAGAGGGTAGAGAAGTGGAGACAATGTGAC
AATATATTGAATAAAGTTGTTTAAAATTTATAACTCTCATAAGTTCATATTATGCTGAAGCTGTAG
CCATCTATAACTGTGTAACATGAAATGTTAAGACATTAACCTAAATACTTCAGCTGATAACAAAAC
AATGTTAATACATACGTCAATGTAACATTTTCTTATCTTTAGGTTATAGCATAAACACTTCAGAGA
TACAGTGACGAAAACCTCTATTTAAATATTTTCAG

DOMAIN 2E

GTTCTTTCTGCGTCCTGATGGGCATTTCAGATGACATCCTTGTGAGAAAAGAAGTGAACAGCCTGA
CAACCAGGGAGACTGCATCTCTGATCCATGCTCTGAAAAGTATGCAGGAAGACCATTACCTGATG
GGTTCCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGCCCTTCACCATCTGCAACTCACC
GTTATGCTTGCTGTGTCCACGGCATGGCTACATTTCCCCAGTGGCACAGACTGTACACTGTACAGT
TCCAGGATGCACTGAGGAGACATGGAGCTGCAGTAGGTGTACCGTATTGGGATTGGCTGCGACCGC
AGTCTCACCTACCAGAGCTTGTACCATGGAGACATACCATGATATTTGGAGTAACAGAGATTTCC
CCAATCCTTTCTACCAAGCCAATATTGAGTTTGAAGGAGAAAACATTACAACAGAGAGAGAAGTCA
TTGCAGACAAACTTTTTGTCAAAGGTGGACACGTTTTTGATAACTGGTTCTTCAAACAAGCCATCC
TAGCGCTTGAGCAGGAAAACACTGTGACTTTGAGATTTCAGTTTGAAATTCTTCACAACGGCGTTC
ACACGTGGGTGCGAGGCAGTCGTACCCACTCTATCGGACATCTCCATTACGCATCCTACGACCCCTC
TTTTCTACCTCCACCATTCCCAGACAGACCGTATTTGGGCAATCTGGCAAGAACTCCAGGAACAGA
GAGGGCTCTCAGGTGATGAGGCTCACTGTGCTCTCGAGCAAATGAGAGAACCATTGAAGCCTTTCA
GCTTCGGCGCTCCTTATAACTTGAATCAGCTAACACAGGATTTCTCCCGACCCGAGGACACCTTCG
ACTACAGGAAGTTTGTTTATGAATATGACAATTTAGAATTCCTAGGAATGTCAGTTGCTGAACTGG
ATCAATACATTATTGAACATCAAGAAAATGATAGAGTATTCGCTGGGTTCTGTTGAGTGGATTCTG
GAGGTTCCGCATCAGTTAATTTCCAGGTTTGTAGAGCTGATTCCACATGTCAGGATGCTGGGTACT
TCACCGTTCTTGGTGGCAGTGTGAGATGGCGTGGGCATTTGACAGGCTATACAAATATGACATTA
CTGAAACTCTGGAGAAAATGCACCTTCGATATGATGATGACTTCACAATCTCTGTCTGAGTCTGACCG
CCAACAACGGAAGTGTCTGAGCAGCAGTCTAATCCCAACACCGAGTGTATATTCAGCGGGGAC
ATC

INTRON 2E/2F-1 (SEQ ID NO:132)

GTAAGTAGTAAACTGCTCAGATTGTTTTTCATAATTACTCCACTATTAAGTAAAAAGTACTAGTAAT
TCAATAGTACTGTTACAGAGAAATGTAACACAATAGACCACAGAGTCCATTTGTTAAACGCCTTT
GGCTTGGTAAGTCTGAGATTTTGGTGACTGATGGAAAGCTAAAATATATTTTGACAG

DOMAIN 2F-1 (1st part of domain f)

GTGACATAAATACCAAGAGCATGTCAGCGAACCGTGTTGCGCGTGAGCTGAGCGATCTGTCTGCGA
GGGACCCGCTCTAGTCTCAAGTCTGCTCTGCGAGACCTACAGGAGGATGATGGCCCCAACGGATACC
AGGCTCTTGACGCCTTCCATGGGCTACCAGCAGGCTGCCATGATAGCCAGGGAAATGAG

INTRON 2F-1/2F-2 (SEQ ID NO:133)

GTATATTTAAGTATTTTATCTTACGCATGACCCTGACCCTATTTATTTTTTTTTTAATCCTCGGATT
TGTTTAAATCCTGTTACCAGCGAAGGTCGGGTTAGAATTGATCTTCAGTCAACTATTCTTGTCGTA
GGACTAACGAGTTGTCTGGCTTGCTTACTCGGTTGACACGTGTCAACGGATCCCAATTGCAATTAG
ATCGATGCTCATGCTGTTGATCCCTGGATTGCCTGGTCCGGACTCCACATACCGCCGCCATATTGC
TGGTATATTGTCGAATGCGACGCTAAACAGCAAGCCAACCAACAATACTGAGACCTGGTGGTACAT
GTCAGTTCTCTATTGCTGGGGTTCCAAACATAGCCATCAGTTGAAATATTTTCATACATAGAAGAAT
ACCTCTGAATATGATGATGAAACATTTACTTAGACTTGCCTGTGAGCCCCAGGCAAAATGCACTGT
AAAAATACACTGACAGAGGATTAGGCATTCTTGGGAGTACTGTATAGTTAGTTGCATACATATTAG
CGTTCCTCCTCACTAAAACGAATCTCTGAATGCTATCAATTAAAGATCATGATGCTTTGATTGTGTCT
ACTGTATTTAAATGGTGTTAAGATTTGCAATTACAATATACACAAACACGTTTCCTGCATCTCGG
AGAATGCAATCTTTCGTTGTACGCGTCTGTTTTCATATTTTATGCATGTAGTTTGCCTACTTAG
CGTCCAATAAATCCATTACAAAATCACACAAACAAACGATTTTAGGAATGTGACTGTAGCTGCAA
CGAATATACCTGATCCTTTCTTGTTCCAG

DOMAIN 2F-2 (2nd part of domain f)

ATCGCATGTTGCATTCACGGTATGCCGACCTTCCCCAGTGGCACAGACTGTACACCCTGCAGTTG
GAGATGGCTCTGAGGAGACATGGATCATCTGTGCGCATCCCCCTACTGGGACTGGACAAAGCCTATC
TCCGAACCTCCCCTCGCTCTTCACCAGCCCTGAGTATTATGACCCATGGCATGATGCTGTGGTAAAC
AACCATTCTCCAAAGGTTTTGTCAAATTTGCAAATACCTACACAGTAAGAGACCCACAGGAGATG
CTGTTCCAGCTTTGTGAACATGGAGAGTCAATCCTCTATGAGCAAACCTCTTCTTGCTCTAGAGCAA
ACCGACTACTGTGATTTTGAGGTACAGTTTGAGGTCCCTCCATAACGTGATCCACTACCTGTTGGC
GGACGTGACACCTACGCATTGTCTTCTCTGCATTATGCATCCTACGACCCATTCTTCTTTATACAC
CATTCTTTGTGGATAAGATGTGGGTAGTATGGCAAGCTCTTCAAAGAGAGGAGGAACTTCCATAC
AAGCGAGCTGACTGTGCTGTCAACCTAATGACTAAACCAATGAGGCCATTTGACTCCGATATGAAT
CAGAACCCATTACAAAGATGCACGCAGTTCCCAACACACTCTATGACTACGAGACACTGTACTAC
AGCTACGATAATCTCGAAATAGGTGGCAGGAATCTCGACCAGCTTCAGGCTGAAATTGACAGAAGC
AGAAGCCACGATCGCGTTTTTGTGATTCTTGCTTCGTGGAATCGGAACCTTCTGCTGATGTCAGG
TTTTGGATTTGTAGAAATGAAAATGACTGCCACAGGGGTGGAATAATTTTCATCTTAGGTGGAGCC
AAGGAAATGCCATGGTCATTTGACAGAACTTCAAGTTTGATATCACCCATGTACTCGAGAAAGCT
GGCATTAGCCCAGAGGACGTGTTTGATGCTGAGGAGCCATTTTATATCAAGGTTGAGATCCATGCT
GTTAACAAGACCATGATACCATCGTCTGTGATCCCAGCCCCAACTATCATCTATTCTCCTGGGGAA
G

INTRON 2F-2/2G-1 (SEQ ID NO:134)

GTGAGAGAACCAGTAATAGCTACTGTCTACAAAGAATGTGTTCAATTTAAAGACCTGACTGTAGGCC
GATGGCTGCTGTCATCTCCTCCGCTCCTCCTCCTGTTCCCTCCGAAGGGGTCAGCTTCAGGTT

CTCTTGCCAATATGCCAAGCAGACCTCCTGAGCAGGCAGTATATATACGTAAGGGAAGCAAGTATG
GACCATCGCGCGGCATGTAGAGATACAATGATCAGCTGTCTGCTGTTCCACTCCTGTCAGACAATG
AGATAAACATGAATACAGTATTACTCAGCAGCGTTCCAATTTTCAACCCTCGTATTTATTAAAAA
AGGAATTTTTAATATATTTTTCTCCTTGTTGAAATATTTTAGTAACTGTTAATCGATATAGAGTGG
AGTAGTGACGCTTTATTTTCGGTTCATTCTCGAAACAAAAATATAATAGTCCACTGAACTCTCTTAA
ATTGTTTTTACAACCTTCAACTGCCACAGACGTAATCCCTCACGTTATTTTGAGCTGACAACGTGT
TGAATTGAGTGTGTTCCGAATTCTAAATAAGCATGTATATATTTACGTCTCATGCAAGTAATATAT
GTTTAACTGATGACGTCACTTGGTGACCACTGATTTAGTTCCTTTGTCATAATTGCAGTTTCTGTT
GTCACGGGGACGGTGGGGAAGCCAGGTTCTCCTGTACGCTGAATATCCCGTTCGAATCCCCCAC
ATGGGTACAAAGTGTGATGCCTATTTCTGGTGTCCCCCACCAGTATATTGCTGGAATAAGTGGCTT
AATACCATATACACTCACTCTATTGTACACTACTGCCACCGGCTCACACCTCTGATGCTTCTGTT
CTATCCAG

DOMAIN 2G-1 (1st part of domain g)

GTCGCGCTGCTGACAGTGCACACTCAGCCAACATTGCTGGCTCTGGGGTGAGGAAGGACGTCACGA
CCCTCACTGTGTCTGAGACCGAGAACCTAAGACAGGCTCTTCAAGGTGTCATCGATGATACTGGTC
CCAATGGTTACCAAGCAATAGCATCCTTCCACGGAAGTCCTCCAATGTGCGAGATGAACGGCCGCA
AGGTTGCCTGTTGTGCTCACG

INTRON 2G-1/2G-2 (SEQ ID NO:135)

GTAATTAATGGATGTGAAGTCAATGTCCGAGGGTATAATAAGGATTTAAATACTTCAGTCGTGTAA
TACTGTATGACATGTGTATTGGATGGTGTAGGTATTACAGGTTATAAGGCCAGTGTGTGTTGGGAC
GGTACTTTTCTGCACTAGTAATAAGCATTGTATTTAGCTAGCTTTTATCATATAACTTTAGTTTC
ATGGTTTGTGGCAATTGAAATCGAAATTTTCTTTTCAAGGTTATCGCACTCGTGTGTTAGAA
TAGTTACTATGCTGCATTGAGAATAACACTATAGTAATAAAGCATATCATAACAGTAAGAATAACAC
TATAGTAATAAAGTATATCATAACAGTAAGAATGTCATTGTATGATAAATAGGTTATCACACTCGTG
TGTTTTAGAAATGGTTACTATCCCAGGAATAACCACTATGTATTACATGTATATTGGGCAGTGTAAG
TAGTAGCATTGTATATTAAATCAGTATATCGTGCTTCAAACACCAGGATATATGGGGTATACAGT
GGGCAGTGTAAGTAGCAACATTGTATATTAAATCAGTATATCGTACTTCAAACACCAGGATTATG
GGGTATACAGTGGGCAGTGTAAGTAGTAGCATTGTATATTAAATCAGTATATCGTACTTCAAACA
CCAGGATATAATTCAGTATATCGTGCTTCAAACACCAGGATATAATTCAGTATATCGTGCTTCAA
AACACCAGGATATATGGGATATACAGTGCGGGTTTGCATACAACCTCCACCCTTTACAG

DOMAIN 2G-2 (2nd part of domain g)

GTATGGCCTCCTTCCCACACTGGCACAGACTGTATGTGAAGCAGATGGAAGACGCCCTGGCTGACC
ACGGATCACATATCGGCATCCCTTACTGGGACTGGACAACCTGCCTTCACAGAGTTACCCGCCCTTG
TCACAGACTCCGAGAACAATCCCTTCCATGAG

INTRON 2G-2/2G-3 (SEQ ID NO:136)

GTCAGTTTAGTCTCCTGTCTGAGCTAACGATACCAATTTCTATTTTCGAGAACCACGATGACGAG
AAAACAAGCAATATAGATATAGATGCAGTATAGATCAAGTTAATGAATTCATTGCTATATGTTTGC
TTGTAATAAACTTTAAGAAAACGAGAGCATGCACACAAATGAAACAAACAATTATGTGTTTGATAG
GAATATGATATATGTATTTGGGGGCTGACGTGAGCAGGGTTGAAGGGACAGTTTACATTGTCAGTA
ACACTGGGAGTATTCTTTGATCCACAATATATAGTTTTCATTGTGTTTACGAGTTACAACCTAACATT
ATATCATACATTACGTGTAACATGCTTCTTTTGTCTCTTCTGCCAG

DOMAIN G-3 (3rd part of domain g)

GGTCGCATTGATCATCTCGGTGTAACCACGTCACGTTCCCCCAGAGACATGCTGTTTAAACGACCCA
GAGCAAGGATCAGAGTCGTTCTTCTATAGACAAGTCCTCCTGGCTTTGGAGCAGACTGACTACTGC

CAGTTCGAAGTCCAGTTTGGAGCTGACCCACAACGCCATTCACTCCTGGACAGGTGGACGTAGCCCT
TACGGAATGTGACCCTCGAGTTCACAGCCTACGATCCTCTCTTCTGGCTTCACCACTCCAACACC
GACAGAATCTGGGCTGTCTGGCAAGCACTGCAGAAATACCGAGGACTCCCATAACAACGAAGCACAC
TGTGAAATCCAGGTCTGAAACAGCCCTTGAGGCCATTCAACGATGACATCAACCACAATCCAATC
ACCAAGACTAATGCCAGGCCTATCGATTCAATTTGATTATGAGAGGTTTAACTATCAGTATGACACC
CTTAGCTTCCATGGTAAGAGCATCCCTGAACTGAATGACCTGCTCGAGGAAAAGAAAAGAGAAGAG
AGAACATTTGCTGCCTTCCTTCTTCGTGGAATCGGTTGCAGTGCTGATGTCGTCTTTGACATCTGC
CGCCCCAATGGTGACTGTGTCTTTGCAGGAACCTTTGCTGTGCTGGGAGGGGAGCTAGAAATGCCT
TGGTCCTTCGACAGACTGTTCCGCTATGACATCACCAGAGTCATGAATCAGCTCCATCTCCAGTAT
GATTCAAGATTTAGTTTCAAGGTGAAGCTTGTGCAACCAATGGCACTGAGCTTTCATCAGACCTC
CTCAAGTCACCAACAATTGAACATGAACTTGGAG

INTRON 2G-3/2H (SEQ ID NO:137)

GTATGTTATCTTATTATCAAATGTGTAATCAGATACTGGAGACGTTTTTCATATTAAGTTGGTCAGC
ATTAGTTGATGATTTTGGTGCGATATTGACGACAAGGAGTTAAGCATTAAACACGTTCAACACATCT
TTAATCTGATATGAGAAGGGAATAAATTGATCCAGTATTGATGATTGAAGTTAGATTAACAGTGAA
AGATATAACAGTTTTTGATAATCGTATAAAACAGTAGCAGAATTGTATCGTGAAAACATAAATGTGGG
AAGGCGAACGCCAAGCAGATTTTAGATTACGATCGTGTGCTAGAATAATTCACAATAACCCAGACG
TCGGAAATGTGGTTGTCTATGGCAATAGTTACGATTAATTGCTAACATGCACGATTTACCTATTTCA
AG

DOMAIN 2H

CCCACAGAGGACCAGTTGAAGAAACAGAAGTCACTCACCAAAATACTGACGGCAATGCACACTTCC
ATCGTAAGGAAGTTGATTCGCTGTCCCTGGATGAAGCAAACAACCTTGAAGAATGCCCTTTACAAGC
TACAGAACGACCACAGTCTAACAGGATACGAAGCAATCTCTGGTTACCATGGATACCCGAATCTGT
GTCCGGAAGAAGGCGATGACAAATACCCCTGCTGCGTCCACGGAATGGCCATCTTCCCCCACTGGC
ACAGACTCTTGACCATCCAACCTGGAAAGAGCTCTCGAGCACAATGGTGCCTGCTTGGTGTTCCTT
ACTGGGACTGGACCAAGGACCTGTCGTCCTGCGGCGTTCTTCTCCGACTCCAGCAACAACAATC
CCTACTTCAAGTACCACATCGCAGGTGTTGGTCACGACACCGTCAGAGAGCCAACTAGTCTTATAT
ATAACCAGCCCCAAATCCATGGTTATGATTATCTCTATTACCTAGCATTGACCACGCTTGAAGAAA
ACAATTACTGTGACTTTGAGGTTCAGTATGAGATCCTCCACAACGCCGTCCACTCCTGGCTTGGAG
GATCCCAGAAGTATTCCATGTCTACCCCTGGAGTATTTCGGCCTTTGACCCTGTCTTTATGATCCTTC
ACTCGGGTCTAGACAGACTTTGGATCATCTGGCAAGAACTTCAGAAGATCAGGAGAAAGCCCTACA
ACTTCGCTAAATGTGCTTATCATATGATGGAAGAGCCACTGGCGCCCTTCAGCTATCCATCTATCA
ACCAGGACGAGTTACCCCGTGCCAACCTCCAAGCCTTCTACAGTTTTTGACAGCCATAAGTTCCGGCT
ACCATTACGATAACCTGAATGTTAGAGGTCACAGCATCCAAGAACTCAACACAATCATCAATGACT
TGAGAAACACAGACAGAATCTACGCAGGATTTGTTTTGTGTCAGGCATCGGTACGTCTGCTAGTGTCA
AGATCTATCTCCGAACAGATGACAATGACGAAGAAGTTGGAACCTTTCACTGTCTGGGAGGAGAGA
GGGAAATGCCATGGGCCTACGAGCGAGTTTTCAAGTATGACATCACAGAGGTTGCAGATAGACTTA
AACTAAGTTATGGGGACACCTTTAACTTCCGACTAGAGATCACATCCTACGATGGATCGGTGGTAA
ACAAGAGCCTACCCAATCCTTTTCATCATCTACAGACCTGCCAATCATGACTACGATGTTCTTGTTA
TCCCAGTAGGAAGAAACCTTTCATCCTCCCAAGTTGTGTCGTCAGAGAGGCACCCGCATCGAGT
TCCACCCAGTCGATGATTCAGTTACGAGACCAAGTTGTTGATCTTGGAAGCTACACTGCACTCTTCA
ACTGTGTGGTACCACCGTTTCACATACCGCGGATTGCAACTGAACCACGTCTATTCTGTCAAGCCTG
GTGACTACTATGTTACCGGACCAACGAGAGACCTTTGCCAGAATGCAGATGTCAGGATTCATATCC
ATGTTGAGGATGAGTAA

3' UTR

CGCAACAG

INTRON 3'UTR (SEQ ID NO:138)

GTGAGATAAGAAACCCTTCTAACAGTAATACGACACCACATTACAGCTTAAACATGATTGCCATCG
ATGTTTTTCATGTGTAGTATACGCTTTTCAGTTCTACATAATTTTGTTTTTCAAATCAAGTTTAGCA
AATGAATCTATCACTGGAAAATAGGGTAGGGTAGCCAAGTGGTTAAAGCGGTCACTGATCACGCCA
AAGACGAGTGTCTTAACCTGCATGGGTACAAAAGTGAAGACCATTGCTGGTGTCTACCGCCGTAAT
ATTGTTTTTAGTATTGCTAAACTTATACTCACCCATGCGCTGTAAAAGTGAATAATAATCATAT
TTCAACAAAAGCACAAAACCATTTCATTTTCATGAAAGCCTCTTGTTACCTGAAAGACGCAAGAG
AACAAATAGTTCCTAACATTATTTTCAGACATTGGAAATGTCCTGCACGTGTAAACCATATATCCTT
TGAAATTTTACGACTGCATCGTATACAATTTATGATATAAATTTAAACTTTATTTTCAG

3'UTR

GTTTCTTGGTCTCCACATATTCACACATCAGCACCAAACGGTTTCGAAGGACATTGGCGTTCTTCT
CTGGCAATGCATTTCAATACAACATTGAAAATGACTTCAGCATATCAGTGTGCTTCGAACGTGTTC
CGGAAGTACTCAAATGTGCTATGACTGAATTATTGTACATACATAACTTATTGATGTTCAATAAAT
AAATGTTGAAACG

Figure 7**Primary structure of the HtH2 protein****DOMAIN A (SEQ ID NO:156)**

GLPYWDWTQHLLTQLPDLVSDPLFVDPEGGKAHDNAWYRGNIKFENKKTARAVDDRLFEEKVGPAGENT
 RLFEGILDALEQDEFNCFEIQFELAHNAIHYLVGGRHTYSMSHLEYTSYDPLFFLHHSNTDRIFAI
 WQRLQVLRGKDPNTADCAHNLIHEPMEPFRSDSNPLDLTRENSKPIDSFDYAHLGYYQYDDLTLNGM
 TPEELNSYLHERSGKEGVFASFRLSGFGGSANVVYACRPAHDEMAVDQCDKAGDFFVLGGPTEMP
 WRFYRAFHFVDVTSIDNIDKDRHGHYYVKAELFSVNGSALPNDLLPQPTISHRPARGHVDEAPAPS
 SDAHLAVRKDINHLTREEVYELRRAMERFQADTSVDGYQATVEYHGLPARCFFPEATNRFACCIHG
 MATFPHW

DOMAIN B

HRLFVTQVEDALIRRGSPIGVPYWDWTQPMALPGLADNATYRDPISGDSRHNPFHDVEVAFENGR
 TERHPDSRLFEQPLFGKHTRLFDSIVYAFEQEDFCDFEVQFEMTHNNIHAWIGGGGKYSMSSLHYT
 AFDPISYLHHSNTDRLWAIWQALQIRRNKPYKAHCAWSEERQPLKPFASFSSPLNNNEKTYENSVPT
 NVYDYEGLVGYTYDDLNFGGMDLGQLEEYIQRQRQDRFTAGFFLSHIGTSANVEIIIDHGTLHTS
 VGTFAVLGGEKEMKWGFDRLYKYEITDELRLNLRRADDGFSISVKVTDVDGSELSSSELIPSAIIIF
 ERSR

DOMAIN C

IDHQDPHQDTIIRKNVDNLTPEEINSLRRAMADLQSDKTAGGFQQIAAFHGEPKWCPSPDAEKKFS
 CCVHGMVFPWHRLTVQGENALRKHGCLGALPYWDWTRPLSHLPDLVSQQNYTDAISTVEARNP
 WYSGHIDTVGVDTTTRSVRQELYEAPGFGHYTGVAQVLLALEQDDFCDFEVQFEIAHNFIHALVGG
 SEPYGMASLRYTTYDPIFYLHHSNTDRLWAIWQALQKYRGKPYNSANCAIASMRKPLQPFGLTDEI
 NPDDETRQHAVPFSVFDYKNNFNIEYDITLDFNGLSISQLDRELSRRKSHDRVFAGFLLHGIQQSAL
 VKFFVCKSDDDCDHYAGEFYILGDEAEMPWGYDRLYKYEITEQLNALDLHIGDRFFIRYEAFDLHG
 TSLGSNIFPKPSVIHDEGA

DOMAIN D

GHHQADEYDEVVTAASHIRKNLKDLSKGEVESLRS AFLQLQNDGVYENIAKFHGKPGLCDDNGRKV
 ACCVHGMPTFPQWHRLYVLQVENALLERGSASVVPYWDWTETFTELPSLIAEATYFNSRQQTDPN
 PFFRGKISFENAVTTRDPQPELYVNRYYYQNVMLAFEQDNYCDFEIQFEMVHNVLHAWLGGRATYS
 ISSLDYSAFDPVFFLHHANTDRLWAIWQELQRYRKPYNEADCAINLMRKPLHPFDNSDLNHDPVT
 FKYSKPTDGFYQNNFGYKYDNLEFNHFSIPRLEEIIIRIRQRQDRVFAGFLLHNIGTSATVEIFVC
 VPTTSGEQNCENKAGTFAVLGGETEMAFHFDRLYRFDISETLRDLGIQLDSDHDFDLSIKIQGVNGS
 YLDPHILPEPSLIFVPGSS

DOMAIN E

SFLRPDGHSDDILVRKEVNSLTRETASLIHALKSMQEDHSPDGFQAIASFHALPPLCPSPSATHR
 YACCVHGMATFPQWHRLYTVQFQDALRRHGAAGVVPYWDWLRPQSHLPPELVMTETYHDIWSNRDFP
 NPFYQANIEFEGENITITEREVIADKLFVKGGHVFDNWFFKQAILALEQENYCDFEIQFEILHNGVH
 TWVGGSRTHSIGHLHYASYDPLFYLHHSQTDRIWAIWQELQEQRLSGDEAHCALEQMREPLKPF
 FGAPYNLNQLTQDFSRPEDTFDYRKFGYEYDNLEFLGMSVAELDQYIIIEHQENDRVFAGFLLSGFG
 GSASVNFQVCRADSTCQDAGYFTVLGGSAAEMAWAFDRLYKYDITETLEKMHRLRYDDDFITISVSLTA
 NNGTVLSSSLIPTPSVIFQRGH

DOMAIN F

RDINTKSMSANRVRRELSDSLARDPSSLKSALRDLQEDDGPNGYQALAAFHGLPAGCHDSQGNEIA
CCIHGMPTFPQWHRLYTLQLEMALRRHGSSVAIPYWDWTKPISELPSLFTSPEYYDPWHDVAVNNP
FSKGFVKFANTYTVRDPQEMLFQLCEHGESILYEQTLLALEQTDYCDFEVQFEVLHNVIHYLVGGR
QTYALSSLHYASYDPFFFIHHSFVDKMVVWQALQKRRKLPYKRADCAVNLMTKPMRPFDSMDNQN
PFTKMHAVPNTLYDYETLYYSYDNLEIGGRNLDQLQAEIDRSRSHDRVFAGFLLRGIGTSADVRFW
ICRNENDCHRGIIIFILGGAKEMPWSFDRNFKFDITHVLEKAGISPEDVFDAAEPFYIKVEIHAVN
KTMIPSSVIPAPTIIYSPGE

DOMAIN G

GRAADSAHSANIAGSGVRKDVTTTLTVSETENLRQALQGVIDDTGPNGYQAIASFHGSPPMCEMNGR
KVACCAHGMAFPHWHRLYVKQMEDALADHGSHIGIPYWDWTTAFTELPALVTDSENNPFHEGRID
HLGVTTSRSPRDMLENDPEQGSSEFFYRQVLLALEQTDYCFEVQFELTHNAIHSWTGGRSPYGMS
TLEFTAYDPLFWLHHSNTDRIWAVWQALQKYRGLPYNEAHCEIQVLKQPLRPFNDNDINHNPIITKTN
ARPIDSFDYERFNYQYDTLSFHGKSIPELNDLLEERKREERTFAAFLLRGIGCSADVVFIDICRPN
DCVFAGTFAVLGGELEMPWSFDRLFRYDITRVMNQLHLQYDSDFSFRVKLVATNGTELSSDLLKSP
TIEHEL

DOMAIN H

GAHRGPVEETEVTHTQNTDGNNAHFHRKEVDSLDEANNLKNALYKLQNDHSLTGYEAIISGYHGYPN
LCPEEGDDKYPCCVHGMAIFPHWHRLTLTQLERALEHNGALLGVYPYWDWTKDLSSLPAFFSDSSNN
NPYFKYHIAGVGHDTVREPTSLIYNQPQIHGYDYLYYLALTTLEENNYCDFEVQYEILHNAVHSWL
GGSQKYSMSTLEYSAFDPVFMIHLSGLDRLWIWQELQKIRRKPYNFAKCAHYHMMEEPLAPFSYPS
INQDEFTRANSPSTVFDSEHKGHYHDNLNVRGHSIQELNTIINDLRNTDRIYAGFVLSGIGTSAS
VKIYLRTDDNDEEVGTFTVLGGEREMPWAYERVFKYDITEVADRLKLSYGDTFNFRLEITSYDGSV
VNKSLPNPFIIRPANHDYDVLVIPVGRNLHIPPKVVVKRGTRIEFHPVDDSVTRPVVDLGSYTAL
FNCVPPFTYRGFELNHVYSVKPGDYVVTGPTRDLCQNADVRIHIHVEDE

Figure 8**Genomic sequence of the KLH1 gene****DOMAIN 1B**

GGCCTACCGTACTGGGACTGGACTGAACCCATGACACACATTCCGGGTCTGGCAGGAAACAAAACCT
TATGTGGATTCTCATGGTGCATCCCACACAAATCCTTTTCATAGTTCAGTGATTGCATTTGAAGAA
AATGCTCCCCACACCAAAAGACAAATAGATCAAAGACTCTTTAAACCCGCTACCTTTGGACACCAC
ACAGACCTGTTCAACCAGATTTTGTATGCCCTTTGAACAAGAAGATTACTGTGACTTTGAAGTCCAA
TTTGAGATTACCCATAACACGATTACACGCTTGGACAGGAGGAAGCGAACATTTCTCAATGTCGTCC
CTACATTACACAGCTTTTCGATCCTTTGTTTTACTTTACCATTTCTAACGTTGATCGTCTTTGGGCC
GTTTGGCAAGCCTTACAGATGAGACGGCATAAACCCCTACAGGGCCCACTGCGCCATATCTCTGGAA
CATATGCATCTGAAACCATTTCGCCTTTTCATCTCCCCTTAAACAATAACGAAAAGACTCATGCCAAT
GCCATGCCAAACAAGATCTACGACTATGAAAATGTCTCCATTACACATACGAAGATTTAACATTT
GGAGGCATCTCTCTGGAAAACATAGAAAAGATGATCCACGAAAACCAGCAAGAAGACAGAATATAT
GCCGTTTTTCTCCTGGCTGGCATACTTTCAGCAAATGTTGATATCTTCATTAAAACTACCGAT
TCCGTGCAACATAAGGCTGGAACATTTGCAGTGCTCGGTGGAAGCAAGGAAATGAAGTGGGGATTT
GATCGCGTTTTCAAGTTTGACATCACGCACGTTTTTGAAAGATCTCGATCTCACTGCTGATGGCGAT
TTCGAAGTTACTGTTGACATCACTGAAGTCGATGGAACATAACTTGCATCCAGTCTTATTCCACAT
GCTTCTGTCATTTCGTGAGCATGCACGTGGTAAGCTGAATAGAG

INTRON 1B/1C (SEQ ID NO:139)

GTTTTGTAATAATTATGTAGAATTCTTTACCTCAGAATAAGATGAGGTCACATGGGTTTTGCAAAA
CTATTACGTTTGAATTAATATTAATAATACCGGACCCTCCACTGGTACATATTTATCTTTATAACG
ATAATAGCGATGATGATGATGATGATGATGATGATGATGATGATAATGATGATGCCGGTATTG
CACGTAATCCAGCCGACTTAGATGACACCCTAAGGGTGCAGAAAGTATAACAATTAGATTGCGTTTT
GCATCTGTGTATGCGTGTGCTTTAACCAAAAGTCAAAATAAAAGTGCAAACCCTTAGTTTTATTCAT
TTGATAGAGCCTTTTACGATAAGAACAATGTAATAAATTAGAACATAACTGAAACCTCCGAAAGAA
GGCCTGTTTGTCAAGAGAGGTATCGACATGATTGACTTATAAACCTGTGCTTCTATATTTTGGAAC
TGTCCACTTTCTTGTGTGTGTACTGTAATCACATCGCACTATGGCTGCAAGACGTGTACGAGTAC
ACTATATACTTACCTAATGACCAACCACAAGGCTGGCTTTGTTAATATTGTTATTTACAGAAATA
AACACAGAATTCAGCATTGCTGGTGTATTTAGCAAAACACCGATATGACACTCATGTTTTATT
ACATTTTTTTTCAG

DOMAIN 1C

TTAAATTTGACAAAGTGCCAAGGAGTCGTCTTATTCGAAAAAATGTAGACCGTTTGAGCCCCGAGG
AGATGAATGAACCTTCGTAAAGCCCTAGCCTTACTGAAAGAGGACAAAAGTGCCGGTGGATTTCAGC
AGCTTGGTGCATTCCATGGGGAGCCAAAATGGTGTCTAGTCCCAGCATCTAAAAAATTTGCCT
GCTGTGTTACGGCATGTCTGTGTTCCCTCACTGGCATCGACTGTTGACGGTTCAGAGTGAAAATG
CTTTGAGACGACATGGCTACGATGGAGCTTTGCCGTACTGGGATTGGACCTCTCCTCTTAATCACC
TTCCCGAACTGGCAGATCATGAGAAGTACGTGACCCTGAAGATGGGGTAGAGAAGCATAACCCTT
GGTTCGATGGTCATATAGATACAGTCGACAAAACAACAAGAAGTGTTTCAGAATAAACTCTTCG
AACAGCCTGAGTTTGGTCATTATACAAGCATTGCCAAACAAGTACTGCTAGCGTTGGAACAGGACA
ATTTCTGTGACTTTGAAATCCAATATGAGATTGCCATAACTACATCCATGCACTTGTAGGAGGCG
CTCAGCCTTATGGTATGGCATCGCTTCGCTACACTGCTTTTGATCCACTATTCTACTTGCATCACT
CTAATACAGATCGTATATGGGCAATATGGCAGGCTTTACAGAAGTACAGAGGAAAACCGTACAACG
TTGCTAACTGTGCTGTTACATCGATGAGAGAACCCTTTGCAACCATTTGGCCTCTCTGCCAATATCA
ACACAGACCATGTAACCAAGGAGCATTCAGTGCCATTCAACGTTTTTGATTACAAGACCAATTTCA
ATTATGAATATGACACTTTGGAATTTAACGGTCTCTCAATCTCTCAGTTGAATAAAAAGCTCGAAG

CGATAAAGAGCCAAGACAGGTTCTTTGCAGGCTTCCTGTTATCTGGTTTCAAGAAATCATCTCTTG
TTAAATTCAATATTTGCACCGATAGCAGCAACTGTCACCCGCTGGAGAGTTTTACCTTCTGGGTG
ATGAAAACGAGATGCCATGGGCATACGATAGAGTCTTCAAATATGACATAACCGAAAACTCCACG
ATCTAAAGCTGCATGCAGAAGACCACTTCTACATTGACTATGAAGTATTTGACCTTAAACCAGCAA
GCCTGGGAAAAGATTTGTTCAAGCAGCCTTCAGTCATTCATGAACCAAGAATAG

INTRON 1C/1D (SEQ ID NO:140)

GTACTTGTTATATGTTTCGAATATTGCCGATACCTTCAATATATATACTTTATCAAAGTAATTGAT
TAATCTGAAGTAATTTTCCTTTCCAGTAGAGATTGAGTTGATACAACAAGAATTCGCCCTGTTGTA
TGTCACCTTTATTTTCATCAAACGATTTCGAAGTGAGCTGTCCATGCCACAATGGGGTCTCTGTAAC
TTCTCGTATGGGGTATAGATTATATAGACGTGGCAGACCTTACGTATACTAATATTTGTGTAATG
TCGTTTTAG

DOMAIN 1D

GTCACCATGAAGGCGAAGTATATCAAGCTGAAGTAACTTCTGCCAACCGTATTCGAAAAACATTG
AAAATCTGAGCCTTGGTGAAGTCTGAGAGCTGCCTTCCTGGAAATTGAAAACGATGGAA
CTTACGAATCAATAGCTAAATTCCATGGTAGCCCTGGTTTGTGCCAGTTAAATGGTAACCCCATCT
CTTGTTGTGTCCATGGCATGCCAACTTTCCCTCACTGGCACAGACTGTACGTGGTTGTCTGTTGAGA
ATGCCCTCCTGAAAAAAGGATCATCTGTAGCTGTTCCCTATTGGGACTGGACAAAACGAATCGAAC
ATTTACCTCACCTGATTTTCAGACGCCACTTACTACAATTCCAGGCAACATCACTATGAGACAAACC
CATTCCATCATGGCAAATCACACACGAGAATGAAATCACTACTAGGGATCCCAAGGACAGCCTCT
TCCATTGAGACTACTTTTACGAGCAGGTCCCTTACGCCTTGGAGCAGGATAACTTCTGTGATTTG
AGATTGAGTTGGAGATATTACACAATGCATTGCATTCTTTACTTGGTGGCAAAGGTAAATATTCCA
TGTCAAACCTTGATTACGCTGCTTTTGATCCTGTGTTCTTCCTTCATCACGCAACGACTGACAGAA
TCTGGGCAATCTGGCAAGACCTTCAGAGGTTCCGAAAACGGCCATACCGAGAAGCGAATTGCGCTA
TCCAATTGATGCACACGCCACTCCAGCCGTTTGATAAGAGCGACAACAATGACGAGGCAACGAAAA
CGCATGCCACTCCACATGATGGTTTTGAATATCAAAACAGCTTTGGTTATGCTTACGATAATCTGG
AACTGAATCACTACTCGATTCCCTCAGCTTGATCACATGCTGCAAGAAAGAAAAAGGCATGACAGAG
TATTCGCTGGCTTCCTCCTTCACAATATTGGAACATCTGCCGATGGCCATGTATTTGTATGTCTCC
CAACTGGGGAACACACGAAGGACTGCAGTCATGAGGCTGGTATGTTCTCCATCTTAGGCGGTCAA
CGGAGATGTCCTTTGTATTTGACAGACTTTACAACTTGACATACTAAAGCCTTGAAAAAGAACG
GTGTGCACCTGCAAGGGGATTTGATCTGGAAATTGAGATTACGGCTGTGAATGGATCTCATCTAG
ACAGTCATGTATCCACTCTCCCACTATACTGTTTGAGGCCGGAACAG

INTRON 1D/1E (SEQ ID NO:141)

GTAATATTTTGTCACTGTAACCAACAAGTGCAGTCTATTTTGAATTACGATAATAACAATTTT
GAAATATATCTTTATTAAAGCAAAGGTTTCTAGAGACAAACAGCCGGCTCTAATTATTTTTCGAA
CTTACGCTTGAGTAAAGATCTGCAAATGGCAACCCTACCTATACTATTAATAATATAATGTTACAT
TCGTATCTGAATGTTTAATAAATCACTTCATATTCTGTTGCAG

DOMAIN 1E

ATTCTGCCCACACAGATGATGGACACACTGAACCAGTGATGATTCGCAAAGATATCACACAATTGG
ACAAGCGTCAACAAGTGTCACTGGTGAAAGCCCTCGAGTCCATGAAAGCCGACCATTTCATCTGATG
GGTTCAGGCAATCGCTTCCTTCCATGCTCTTCCTCCTCTTTGTCCATCACCAGCTGCTTCAAAGA
GGTTTGCGTGCTGCGTCCATGGCATGGCAACGTTCCCAACAATGGCACCCTGTGTACACAGTCCAAT
TCCAAGATTCTCTCAGAAAAATGGTGCAGTCGTTGGACTTCCGTACTGGGACTGGACCCTACCTC
GTTCTGAATTACCAGAGCTCCTGACCGTCTCAACTATTTCATGACCCGGAGACAGGCAGAGATATAC
CAAATCCATTTATTGGTTCTAAAATAGAGTTTGAAGGAGAAAACGTACATACTAAAAGAGATATCA
ATAGGGATCGTCTCTCCAGGGATCAACAAAAACACATCATAACTGGTTTATTGAGCAAGCACTGC
TTGCTCTTGAACAAACCAACTACTGCGACTTCGAGGTTGAGTTTGAATATATGCATAATGGTGTTT

ATACCTGGGTTGGAGGCAAGGAGCCCTATGGAATTGGCCATCTGCATTATGCTTCCTATGATCCAC
TTTTCTACATCCATCACTCCCAAACCTGATCGTATTTGGGCTATATGGCAATCGTTGCAGCGTTTCA
GAGGACTTTCTGGATCTGAGGCTAACTGTGCTGTAAATCTCATGAAAACCTCCTCTGAAGCCTTTCA
GCTTTGGAGCACCATATAATCTTAATGATCACACGCATGATTTCTCAAAGCCTGAAGATACATTCTG
ACTACCAAAAAGTTTGGATACATATATGACACTCTGGAATTTGCAGGGTGGTCAATTTCGTGGCATTG
ACCATATTGTCCGTAACAGGCAGGAACATTCAAGGGTCTTTGCCGGATTCTTGCTTGAAGGATTTG
GCACCTCTGCCACTGTCGATTTCCAGGTCTGTGCGACAGCGGGAGACTGTGAAGATGCAGGGTACT
TCACCGTGTGGGAGGTGAAAAAGAAATGCCTTGGGCCTTTGATCGGCTTTACAAGTACGACATAA
CAGAAACCTTAGACAAGATGAACCTTCGACATGACGAAATCTTCCAGATTGAAGTAACCATTACAT
CCTACGATGGAACGTACTCGATAGTGGCCTTATTTCCACACCGTCAATCATCTATGATCCTGCTC
ATC

INTRON 1E/1F (SEQ ID NO:142)

GTAAGTATACACACATTATTTCTCTTCTGCTATATCAGATGAAGAGAACGTTGTATCACTAACCTA
GTCTTGTTTGATTTGTGGTTTCGTTTGCTTCCTGAACAGTAGGGTTGATTTAACTTCTCTGTTTCG
TCTGTACCAATGAAAGACTATGATGCTTGTGTGAAGATGCTTTGTTTCATGAGTCAGTCTGTTCTTG
TAATGCTTTGATCTTTGCCATCAACATTCTTGAAATTAATTATGGTTTCCCTTAAATACTTACATA
TTACATTTAAACGTCGCTGCTTGTCTGATTGCATATTCTTTCAAAAATAACTATATATTCCAG

DOMAIN 1F-1 (1st part of domain f)

ATGATATTAGTTTCGCACCACCTGTCGCTCAACAAGGTTTCGTCATGATCTGAGTACACTGAGTGAGC
GAGATATTGGAAGCCTTAAATATGCTTTGAGCAGCTTGCAGGCAGATACCTCAGCAGATGGTTTTG
CTGCCATTGCATCCTTCCATGGTCTGCCTGCCAAATGTAATGACAGCCACAATAACGAG

INTRON 1F-1/1F-2 (SEQ ID NO:143)

GTAAATATACAGTGAAATCCGGATAAGTAAATCCAGATAAGAAAAAAACATTTTCTGTGGTCCC
GGCATGTTTCTTCTTCATCTATCATTATTTTGATACGGATAAGTAAAAATCGGCTGAGTAAACAT
CCGGGTAAGTAAATGATTTTCGAGGTCTCTTCATCGGATAAGTAAGATACACAAGTGATCATTCC
AATAAACACTAACTGATGCAACACAATACCAGCGCACAGTGTTCCTACTACGTTTGTTTGTATTGT
AATTAACAATTAACACTTAAGTGTTCCTCAATGTGTCCGTGTGCAAACCTGATTGGGACAAAGCTTG
CAACAAGCCCCGCAATTCCATGTCGTTTATGTCTACGTTTGTTATTCTGACTGCTTGGAGGGGTTT
GGAAAAAATAAAAAACGGGTAAATATTATAAAAAATTCACGGTGCCTTGAAATTTTAGGTGTCCG
GATTTCACTGTAGATGATTAATTTCTCACTTGTAACAAAAGGACCCAGTACCCTCATTTCGTGAC
GTACGTTATAAAATGTAATTATAAAAAAGCCATTATCATGTTATACGTGATCTTGNCTTGCAATTA
TNCTACCGCTTTCTTGATTTTTTAAAGCAATTTCTCCCTCTATGAACCTATTAACATAGCACTCCT
GCAAAAGAAAACAGTCACTGCATGGATCCATATTGAATGTTGCTGCTTATTTCTCATTTTATTACT
CACAGATATTTCAAGAACATCGTACTCTTAACCAGGCTAAAGCAAAGAGGGTTACATTTTAGCCG
ACAAGTTCACTAGCTGAGTGGAACACGTATATATTAATGGAGATGACTCTGGTCATGATGATTAGG
ACAATTATCATGACGTTATCATTGATCATGACCATGTGAGTATAATAGATAGCTAACAAATAATGT
AATTACTAATTATGAAGCAATGGTGCATTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

GTGGCATGCTGTATCCATGGAATGCCTACATTCCCCCACTGGCACAGACTCTACACCCTCCAATTT
GAGCAAGCTCTAAGAAGACATGGCTCTAGTGTAGCAGTACCCTACTGGGACTGGACAAAGCCAATA
CATAATATTCCACATCTGTTTACAGACAAAGAATACTACGATGTCTGGAGAAATAAAGTAATGCCA
AATCCATTTGCCCGAGGGTATGTCCCTCACACGATACATACACGGTAAGAGACGTCCAAGAAGGC
CTGTTCCACCTGACATCAACGGGTGAACACTCAGCGCTTCTGAATCAAGCTCTTTTGGCGCTGGAA
CAGCACGACTACTGCGATTTTGCAGTCCAGTTTGAAGTCATGCACAACACAATCCATTACCTAGTG
GGAGGACCTCAAGTCTATTCTTTGTCATCCCTTCATTATGCTTCATATGATCCGATCTTCTTCATA
CACCCTCCTTTGTAGACAAGGTTTGGGCTGTCTGGCAGGCTCTTCAAGAAAAGAGAGGCCTTCCA

TCAGACCGTGCTGACTGCGCTGTTAGTCTGATGACTCAGAACATGAGGCCTTTCATTACGAAATT
AACCATAACCAGTTCACCAAGAAACATGCAGTTCCAAATGATGTTTTCAAGTACGAACTCCTGGGT
TACAGATACGACAATCTGGAAATCGGTGGCATGAATTTGCATGAAATTGAAAAGGAAATCAAAGAC
AAACAGCACCATGTGAGAGTGTTCAGGGTTCCTCCTTACGGAATTAGAACCTCAGCTGATGTC
CAATTCCAGATTTGTAAAACATCAGAAGATTGTACCATGGAGGCCAAATCTTCGTTCTTGGGGG
ACTAAAGAGATGGCCTGGGCTTATAACCGTTTATTCAAGTACGATATTACCCATGCTCTTCATGAC
GCACACATCACTCCAGAAGACGTATTCATCCCTCTGAACCATTCTTCATCAAGGTGTCAGTGACA
GCCGTCAACGGAACAGTTCTTCCGGCTTCAATCCTGCATGCACCAACCATTATCTATGAACCTGGT
CTCGGTG

INTRON 1F-2/1G-1 (SEQ ID NO:144)

GTCTCGGTGAGTTATTAAGAAACAAAATATTTACCATTACCATTGTAACTACAAAATGAGTG
AGATATCTTATATCACTGGTACACTACTGATATTTTATGCAATGAAATTACTATTTTTCCAGGTAC
GCTTCAACCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCATCATGCTTTTCTGT
AAAACATAAAACACCAATTAACAATGTTCTTAGTGTGTTTGTGACTCCCTTCCACTGCAACGCCT
ACATAATCAAAGTGTTCTTTTTTCCAACTTTCCAGTTAGTGTTGAAGACTAAAAGTTAAATA
AGCATTACATAACTTCTAAGAGCAACTGGGACCATGCAGTTACGTATTGATATTTCTGTGAGAGT
GAAGCAAAACACTGTTTTTCAAGCTTAGGTTTATCAATCAAATGTCCAATAGTTCATGTTATCGA
AAAGGCAGCGAAGGATAAGAGGCTCCGAGACATCTTGTCTATTCTCGTGTTTCATATGATATCAACT
GAGGAGCTTCCATTACATTTTTGACCTTATCATTTAAAGACATACATGGAACATTTTCATTTTACA
GTTAAAGTGAACCACTTCAGGTTCAACTTCAACTTCGAATTCAACTTCTGTTGTGTGTTTTATGAG
CCGACTGAAATAGAGTGCCTTACTTTCACCTCTAGTTTCGTTCTGTCTCGTCATCGTTGTTTCTTT
CAGTGTGCATAGTACACGCCTAGTATAGAACACACGAACCTGTCCTTACTTAATAGATTCTGAAAC
TATTATGTGGAAAGTTGGCAGGCTATAGTAACATCCTGGCAAAATTATCATGTATCCTCTTGTTTG
TCATAATTAG

DOMAIN 1G-1 (1st part of domain g)

ACCATCACGAAGATCATCATTTCTTCTTCTATGGCTGGACATGGTGTGAGAAAGGAAATCAACACAC
TTACCACTGCAGAGGTGGACAATCTCAAAGATGCCATGAGAGCCGTCATGGCAGACCACGGTCCAA
ATGGATAACCAGGCTATAGCAGCGTTCCATGGAAACCCACCAATGTGCCCTATGCCAGATGGAAAGA
ATTACTCGTGTTGTACACATG

INTRON G1-1/1G-2 (SEQ ID NO:145)

GTATGTATTTCCCACTGGTGGTCGCTGACTGCCAACACATACTTGTAATTTATTCATGAAAGTATA
ATAGTTTGTGTTGAAAGTATATTTATAACCATCTTGACAAGCGTCACGAATTTTCACCACAAAGCT
TCAAAACGCCCAAAACATTCTAATAGCGATATATTTGTTAAAAGACCAAAATATAGCCTTACAACA
ATAGATTATTTTAATAAGACCAGTCAGTGCATGCAATCGATTGGAACTTTGAAATAAAATATTC
TATGTACTAACTGCCAATCTCATAAATACTTGCCTTGGATGTGCTTCTTTTTTACATTTCGCGTCGAG
CTTCAACTCCAATGCATAAGCTTAAAAATAATCATAAACACAAACAAATAGCCACAGAGGCGACGA
TCCCTCCAGGCCAGGCTTTATTTGTCTCTTATAGAATATATCGCTATTAGAATGTTTTTGACGTTT
TGAAGCTTTGTGGGTGAAATTCGTGATGTTTATGCGTGGTATTTATGTAAGATGAAATAAATAT
ATCTTTTCAAACAAGATTTTAGTATTTTGAAGACTTCTATGAATAAATTACACTTATGTGTTAGGT
TATTGGTCACTGAGCGCTTGTGGTATTTTCCCTTCTTCAATTTGTTTGTCTTTGTTCAATTTTCA
ATAGTTATCCTACTGTGGATAGTCTATATGAGAATCGTTGAAAGAATAATACAATTCTAATGGATT
GCAACTTCTTTAACTTTTATTTGCAACTGCCACGTTTCGGTATACGTTCTTATGCCGTCATCAAGC
ATACGAGTGTACATGTATGCCAAAACGCTGCAATAAAAAATTAAAGAAGTTGCAATCCATAAGAAT
TTCAATGTTCTTTTCATCATCACATCAACTTCTAAAAATGCCTATAAAACAATCAACAAACGTACAA
TAGTACATTACCGGATCTCGCAGCATGACCACGTCGATATCTAAACAATATCACTATCCATTAAATA
GGATCAAGAGTAGGTACAGACATGTTTCAGTTATAAATACTCTTCAAAAAAGTAGGGGAACCTTGAA
TTTCAAGGTCAATAACAACTAATGATAATAACAATTGGTCCCAATAATAACAATTGGTCCCAAA
CTAATTGTATCTTTACAAAGAAGAAATTGAGTGAACAATTCACCCGGTATTTTATTACCTAAACCG

TTTCTCTTGCTGTTATGGTGCGTGAAAGAAGAAATGGGTAAGAAACGGAAATTGACATTTTTGCGT
 CAGTGGTGCGTAATGCCCCATTGTTGGCCAAACACTGATTGATTGCTGAGGCATCGTGCATACG
 CGTCTACCTATGGTAATTTGATGCAGTCTGTCCCATTTCTCCACCAACGCCTGGACAAGTTCATCT
 AGCGTGGCTGGTGGCCTTTTCACGTTGACGCACACGTCGGCCCAAGATGTCCCAGACATTTTCAATG
 GCCAGGGCTCATTGCTGGTCAGGGCATCCTATGGATATTGTGCCGTTGAAGGTGGTTATGTTGTTT
 ACATTGAAATTCCAAGTTCTCCTACTCTTTTTAAGAGGAGGTTACAAAAGTACGTTCTTTTCATGTT
 GGTGAAGAGAATATCAAGGTCTTCTAAGGGATTGTGTCTTATAATATTTGATTTTAAGAAGTTTGA
 TATTATCTGCATCCTTCCCAAGAAATTGCAAATGTTTACACACTATTGCGTTTGATAATGTTTTTG
 GGGAAATAAACTGTCCAGGACTGCTAAATAGTAATTATTGCTACTTTTAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACTTTCCCCCACTGGCACAGACTGTACACAAAACAGATGGAAGATGCCTTGACCGCCC
 ATGGTGCCAGAGTCGGCCTTCCTTACTGGGACGGGACAACCTGCCTTTACAGCTTTGCCAACTTTTG
 TCACAGATGAAGAGGACAATCCTTTCCATCAT

INTRON 1G-2/1G-3 (SEQ ID NO:146)

GTGAGTTCACGTAAGCCTACGAGATCAACATTACTCCTTAACAGCCACGGCATCATGTACCGATAT
 ATCACAAACAAAAGTATTCAAAGCTTTAAACACGATATGTATGGTTCAAGAATGACATCATTAAC
 AAGGACATGAGTCTGAAATAAACATGACTTGACACCGTTGTGGTCACAGTTTTGTTTCTCATTGGT
 GAACCTGTGAAACAACCTTTCAAACCAAAAGATGCCTATTAATATTGTTAATTTCCATGAATTAGG
 AGATACACACATTCTACTGTCATTT.....AATAACCGCTTC
 CAGCATGAAAACACAATATGATTATCTCAATTCTACCATTACTAATTATAATTTTACTGCGCATTA
 TTTGACGACGCGTAAACATCGCTGCTTTACAGACTGCACTGCGGTAACGTGACGTTTTTCATGAC
 GTCACTACATTCTATTCAAACATTTCCACAGAAGAGCGAGACCACGGCCGTGATGGGTTCTGGGC
 AGATGATTACCCAAGTATATATTTATAATAACTTGACTGCTTGCTGAATAATGTTGACACATGAC
 AACGAATTTGTGATAGCGTAAGAAGCGTGAATACTGTGAATAGTGTGAGGGGTGTTTGCTGAGAGT
 TAACCACCGTTAATTGCAAAATTCCCGAATACTTGCAATTTGCAGTCGAAGAAGAATTGCATTCTTA
 CTCCTGTGAATGGACTCATTGTTATTTAGCAGCGGTTATTGAGGTTTTGATCACCTCTAAATAGAC
 AATCAGGATGCGGCAAACCGGAAAATTATAGCAGAATCTGTAATTCAAGATGGGCTTGCCTGTGAA
 AATATGCTGCGAGTTCAGTAACACTTTTCCCTTTTCGATCATGGCCTGTTTTGCTCTGAATCTGGTC
 TTTTCAAGAGGATCCCTGCTTTTTTAAACTAAAGTCCTCCCAACTCACTTATATTTATGTTTTTTAA
 TTATTTATAGTTTTAATATGAACAACAAATCATATTTATTTACACATTATATTTTTTTCAG

DOMAIN 1G-3 (3rd part of domain g)

GGTCACATAGACTATTTGGGAGTGGATACAACCTCGGTCGCCCCGAGACAAGTTGTTCAATGATCCA
 GAGCGAGGATCAGAATCGTTCTTCTACAGGCAGGTTCTCTTGGCTTTGGAGCAGACAGAT

Figure 9

Primary structure of the KLH1 protein

DOMAIN B

GLPYWDWTEPMTHIPGLAGNKTYVDSHGASHTNPFHSSVIAFEENAPHTKRQIDQRLFKPATFGHH
TDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSMSSLHYTAFDPLFYFHHSNVDRLWA
VWQALQMRRHKPYRAHCAISLEHMHLPFAFSSPLNNEKTHANAMPNKIYDYENVLHYTYEDLTF
GGISLENIEKMIHENQQEDRIYAGFLLAGIRTSANVDIFIKTTSVQHKAGTFAVLGGSKEMKWGF
DRVFKFDITHVLKDLDLTADGD FEVTVDITEVDGTKLASSLIPHASVIREHARGKLN

DOMAIN C

VKFDKVPRSLIRKNVDRLSPEEMNELRKALALLKEDKSAGGFQQLGAFHGEPKWCPSPPEASKKFA
CCVHGMSVFPWHRLTLVQSENALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNP
WFDGHIDTVDKTTTRSVQNKLFEOPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGG
AQPYGMASLRYTAFDPLFYLHHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMPREPLQPFGLSANI
NTDHVTKEHSVPFNVFDYKTNFNYEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL
VKFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFYIDYEVFDLKP
SLGKDLFKQPSVIHEPRI

DOMAIN D

GHHEGEVYQAEVTSANRIRKNIENLSLGELESRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPI
SCCVHGMPTFPHWHRLYVVVENALLKKGSSVAVPYWDWTKRIEHLPHLISDATYNSRQHHYETN
PFHHGKIITHENEITTRDPKDSL FHSDFYEQVLYALEQDNFCDFEIQLEILHNALHSLGKGKYS
MSNLDYAAFDPVFFLHHATTDRIWAIWQDLQRFKRKPYREANCAIQLMHTPLQPFDKSDNNDEATK
THATPHDGFYQNSFGYAYDNLELNHYSIPQLDHMLQERKRHRVDFAGFLLHNIGTSADGHVFCVCL
PTGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVHLQGD FDL EIEITAVNGSHL
DSHVIHSPTILFEAG

DOMAIN E

TDSAHTDDGHTPEVMIRKDITQLDKRQQLSLVKALESMKADHSSDGFQAIASFHALPPLCPSPAAS
KRFACCVHGMATFPQWHRLYTVQFQDSL RKHGAVVGLPYWDWTLPRSELPELLTVSTIHD PETGRD
IPNPFIGSKIEFEGENVHTKRDINRDRLFQGSTKTHHNWFIEQALLALEQTN YCDFEVQFEIMHNG
VHTWVGKPEPYGIGHLHYASYDPLFYIHHSQTDRIWAIWQSLQRFRLSGSEANCAVNLMKTPLKP
FSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGIDHIVNRNQEHSRVFAGFLLEG
FGTSATVDFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQIEVTI
TSYDGTVLDSGLIPTPSIIYDPAH

DOMAIN F

HDISSHHLNLNKRHDLSTLSERDIGSLKYALSSLQADTSADGFAAIASFHGLPAKCND SHNNEVA
CCIHGMPFPHWHRLYTLQFEQALRRHGSSVAVPYWDWTKPIHNIPLFTDKEYYDVWRNKVMPNP
FARGYVPSHDTYTVRDVQEGFLHFLTSTGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGG
PQVYSLSSLHYASYDPIFFIIHHSFVDKVWAVWQALQEKRLPSDRADCAVSLMTQNMRFHYEINH
NQFTKKHAVPNDFVKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHRVVFAGFLLHGIRTSADVQF
QICKTSEDCHHGGQIFVLGGTKEMAWAYNRLF KYDITHALHDAHITPEDVFHPSEPPFFIKVSVTAV
NGTVLPASILHAPTIIYEPGLG

DOMAIN G

DHHEDHHSSSMAGHGVRKEINTLTAEVDNLKDAMRAVMADHGPNQYQAIAAFHGNPPMCPMPDGK
NYSCCTHGMATFPHWHRLYTKQMEDALTAHGARVGLPYWDGTTAFTALPTFVTDEEDNPFHHGHID
YLGVDTTSPRDKLFNDPERGSESFFYRQVLLALEQTD

Figure 10**Genomic sequence of the KLH2 gene****DOMAIN 2B**

GGCCTGCCCTACTGGGATTGGACCATGCCAATGAGTCATTTGCCAGAACTGGCTACAAGTGAGACC
TACCTCGATCCAGTTACTGGGGAACTAAAAACAACCCTTTCCATCACGCCCCAAGTGGCGTTTGAA
AATGGTGTAAACAAGCAGGAATCCTGATGCCAACTTTTTATGAAACCAACTTACGGAGACCACACT
TACCTCTTCGACAGCATGATCTACGCATTTGAGCAGGAAGACTTCTGCGACTTTGAAGTCCAATAT
GAGCTCACGCATAATGCAATACATGCATGGGTTGGAGGCAGTGAAAAGTATTCAATGTCTTCTCTT
CACTACACTGCTTTTGATCCTATATTTTACCTCCATCACTCAAATGTTGATCGTCTCTGGGCCATT
TGGCAAGCTCTTCAAATCAGGAGAGGCAAGTCTTACAAGGCCCACTGCGCCTCGTCTCAAGAAAGA
GAACCATTAAGCCTTTTGCAATTCAGTTCCCCACTGAACAACAACGAGAAAACGTACCACAACCTCT
GTCCCCACTAACGTTTATGACTATGTGGGAGTTTTGCACTATCGATATGATGACCTTCAGTTTGGC
GGTATGACCATGTCAGAACTTGAGGAATATATTCACAAGCAGACACAACATGATAGAACCTTTGCA
GGATTCTTCTTTTCATATATTGGAACATCAGCAAGCGTAGATATCTTCATCAATCGAGAAGGTCAT
GATAAATACAAAGTGGGAAGTTTTGTAGTACTTGGTGGATCCAAAGAAATGAAATGGGGCTTTGAT
AGAATGTACAAGTATGAGATCACTGAGGCTCTGAAGACGCTGAATGTTGCAGTGGATGATGGGTTC
AGCATTACTGTTGAGATCACCGATGTTGATGGATCTCCCCCATCTGCAGATCTCATTCCACCTCCT
GCTATAATCTTTGACGTGGTCAGAG

INTRON 2B/2C (SEQ ID NO:147)

GTATTTAAAAAAGTAATAAAACCATATTTTCGAATGCGCTTTATGAAATATCGTGTGACTGGTTCT
TTAGTTTACATGGAGTGTAACAACATGCTCCATCAGTTGACATATACTGCTCACACAAAGTAAGGG
ATATTTGATAATGATAACAAATATAATCAAAGCGGTTATACTATCAAGACTTATTCACATAATTAC
AGGTGAAGGGAGGTGTGATCGTGTTCCTGATCAGGTTGAGGCCAGAGAAGTCCCAGTTTGAGTCT
TGCAGAAGATGATGTTTAGGCATGGGGTCTGAATCACCAAAATCACATGACTTCAATAACGGGTG
ACCACCTCGAGCGACGATGCAAGCAGTAGAGCGTCTACGCATGCTCCTGATAAGGCGACCAATCTG
TTCTTGGGGAATCAGTCGCCACTCCTCTTGTAGTGCCACGCTCATTCTGCTACGGTCTGGGTAC
CTGCTATCGGGTCTTGATCCGTATCCCAAGGATGTCCACACATGTTCAAGGTGAGAGGTCTGGGGA
ACATCGCTGGCCACGGTAAGGTCTGAATTTGATGCCGTTGAAAGTGAGCTCTGACAACCTGAGCAT
GGTGAGCTCTGACGTTGTCTGCTGAAAGATGAATCCAGCTCCATGACAGCGAGCAAAGGGCAGGA
CGTGTGGTCAATGCAGTTGTCTCTGCAGTACACACCTGTCACTCGCCACTCACAAGCGTGTAGAT
CTGTACGACCAGTCATGGAGATCCCAGCCCACATCATAACGGACCCCTATCCATACCGATCATGAG
CCACCATAGCAGCGTCTTGATGACGTTCTCCCTGTGCGCTCGACATCCTCACACGGCCAAAAGGAA
CGTGGACTCGTCACTGAACATGACATTAGCCAACCTGGCACTTGTCCACCGCTGATGTTGGCGAGA
CCATTCCAGTCGAGCTCTTCGGTGTCTGGCTTTTCATCGATAACACGACGTAAGGTCTGCGGGCGTG
CAAGACGGCTCTATGCAGGCGATTTCCGATTGTCTGGGTGCTAACTCTGATCCCAGGTGCCTGCTG
AAGTTGATGCTGGATCTGTGTGGCATTGAGATGGCGATTCTTAGGACTGTGGAGATGATGAATCG
ATCTTGACTTATGGTGGTGACATTAGGACGTCGGGTCTGTGCTCCTATCCTGCACTCTTCCAGTTGT
TCGGTGACGCTCTGGTACCCGGCTGATTACTGACTGAGAATATCCATCTGCCGTGCGACATGAGCC
TGTGTTGGCCCAGCCTGAAGCATTGCAATCGCCAGAGACGCTCTTCAAAGTCAATTCGACGCATGG
TTTTCTGTTCAAAATGACAGCGTAAAACAGTTTTTTGGTGCTTTTATGCTTCCCAAGAGCATGAAA
AACACGTTCTATGGGTCTGTGCACACCTTACATGACAAGTGTGAAAAGTGACTTGCACCCCTTGTG
TGTTTCGGATGCACACTCTGTTTACGTACTGATGCGATTTGGCGTCTAAACATGTTTTGGCGTCTAA
ACATGTTTTCTGTCATGATTCATATACTATTTTGTATATTCTTGGCATCAAACCAAACCTACAGTG
AAATATATTTCAATATCCCTACTTTGTGTGAGTAGTATAGATCACTGCAGACAACATATAGACAA
TGCAGTTACACCGTCAACAATCCAGTCATTAATTATGATGACACTTCCACACATAGTGTCAGTGA
TTGTAATTCAACTGTACACACTTTTCCCGTGAACATTCAGGATCTATATGACTAAATATATAACAT
TAGTATACGTGCAGTTTTGTATCGCTACGACATTGTTGTAACCTTTGTTTAATCATTTAACAG

DOMAIN 2C

CTGATGCCAAAGACTTTGGCCATAGCAGAAAAATCAGGAAAGCCGTTGATTCTCTGACAGTCGAAG
AACAAACTTCGTTGAGGCGAGCTATGGCAGATCTACAGGACGACAAAACATCAGGGGGTTTCCAGC
AGATTGCAGCATTCCACGGAGAACC AAAATGGTGTCCAAGCCCCGAAGCGGAGAAAAAATTTGCAT
GCTGTGTTTCATGGAATGGCTGTTTTCCCTCACTGGCACAGATTGCTGACAGTTCAAGGAGAAAAATG
CTCTGAGGAAACATGGATTTACTGGTGGATTGCCCTATTGGGACTGGACTCGGCCAATGAGCGCCC
TTCCACATTTTGTGCTGATCCTACTTACAATGATTCTGTTTCCAGCCTCGAAGAAGATAACCCAT
GGTATCATGGTCACATAGATTCTGTTGGGCATGATACTACAAGAGCTGTGCGTGATGATCTTTATC
AATCTCCTGGTTTTCGGTCACTACACAGATATTGCAAAACAAGTCCTTCTGGCCTTTGAGCAGGACG
ATTTCTGTGATTTTGAGGTACAATTTGAAATTTGCCATAATTTTCATACATGCTCTGGTTGGTGGTA
ACGAACCATAACAGTATGTCATCTTTGAGGTATACTACATACGATCCAATCTTCTTCTTGCAACCGCT
CCAATACAGACCGACTTTTGGGCCATTTGGCAAGCTTTGCAAAAATACCGGGGGAAACCATAACA
CTGCAAACTGTGCCATTGCATCCATGAGAAAACCACTTCAGCCATTTGGTCTTGATAGTGTCTATA
ATCCAGATGACGAAACTCGTGAACATTCCGTTCCCTTCCGAGTCTTCGACTACAAGAACAACCTCG
ACTATGAGTATGAGAGCCTGGCATTTAATGGTCTGTCTATTGCCCAACTGGACCGAGAGTTGCAGA
GAAGAAAGTCACATGACAGAGTCTTTGCAGGATTCCCTTCTTCATGAAATTGGACAGTCTGCACTCG
TGAAATTCTACGTTTGCAACACAATGTATCTGACTGTGACCATTATGCTGGAGAATTCTACATTT
TGGGAGATGAAGCTGAGATGCCTTGGAGGTATGACCGTGTGTACAAGTACGAGATAACACAGCAGC
TGCACGATTTAGATCTACATGTTGGAGATAATTTCTTCCTTAAATATGAAGCCTTTGATCTGAATG
GCGGAAGTCTTGGTGGAAGTATCTTTTCTCAGCCTTCGGTGATTTTTCGAGCCAGCTGCAG

INTRON 2C/2D (SEQ ID NO:148)

GTATGTTTTAAATGTCACTTATCCGTGATCTGTAATGAAGTTAGCAATTCACCTTTATCAACTGTTT
GGCTGTACTGTTTTCAGTGCGAGTTTTACTTAGGTTGGATTAAATTAATAATTCAGCTCATAAATG
TTTTGATTCAACTTTTGTTATTTATTTCAAACAG

DOMAIN 2D

GTTACACCAGGCTGATGAATATCGTGAGGCAGTAACAAGCGCTAGCCACATAAGAAAAAATATCC
GGGACCTCTCAGAGGGAGAAATTGAGAGCATCAGATCTGCTTTCCTCCAAATTCAAAAAGAGGGTA
TATATGAAAACATTGCAAAGTTCCATGGAAAACCAGGACTTTGTGAACATGATGGACATCCTGTTG
CTTGTTGTGTCCATGGCATGCCACCTTTCCCCACTGGCACAGACTGTACGTTCTTCAGGTGGAGA
ATGCGCTCTTAGAACGAGGGTCTGCAGTTGCTGTTCCCTTACTGGGACTGGACCGAGAAAGCTGACT
CTCTGCCATCATTAAATCAATGATGCAACTTATTTCAATTCACGATCCCAGACCTTTGATCCTAATC
CTTTCTTCAGGGGACATATTGCCTTCGAGAATGCTGTGACGTCCAGAGATCCTCAGCCAGA ACTAT
GGGACAATAAGGACTTCTACGAGAATGTCATGCTGGCTCTTGAGCAAGACA ACTTCTGTGACTTTG
AGATTACGCTTGAGCTGATACACAACGCCCTTCATTCTAGACTTGGAGGAAGGGCTAAATACTCCC
TTTCGTCTCTTGATTATACCGCATTTGATCCTGTATTTTTCTTCACCATGCAAACGTTGACAGAA
TCTGGGCCATCTGGCAGGACTTGACAGAGATATAGAAAGAAACCATAACAATGAGGCTGACTGCGCAG
TCAACGAGATGCGTAAACCTCTTCAACCATTTAATAAACCAGAACTTAACAGTGATTCCATGACGC
TTAAACACAACCTCCACAAGACAGTTTTGATTATCAAAACCGCTTCAGGTACCAATATGATAACC
TTCAATTTAACC ACTTCAGCATACAAAAGCTAGACCAA ACTTATTCAGGCTAGAAAACAACACGACA
GAGTTTTTGTGGCTTTATTCTTCACAACATTGGGACATCTGCTGTTGTAGATATTTATATTTGCG
TTGAACAAGGAGGAGAACAAA ACTTGCAAGACAAAGGCGGGTTCTTCACGATTCTGGGGGGAGAAA
CAGAAATGCCATTCCACTTTGACCGCTTGACAAATTTGACATAACGCTGCTCTGCATAAACTTG
GTGTTCCCTTGGACGGACATGGATTGACATCAAAGTTGACGTCAGAGCTGTCAATGGATCGCATC
TTGATCAACACATCCTCAACGAACCGAGTCTGCTTTTTGTTCCCTGGTGAACGTAAGAATATATATT
ATG

INTRON 2D/2E (SEQ ID NO:149)

GTTATAAAGCAGTATATTCTCTTCAAAAAAGTAGGGGAACTTGGAATTTCAAGGTAAATAACATAA
CTACCTTCAACGGCACAATATCCATATGATGCCCTGGCCAGCAATGAGGCCTGATCTTTTCCCAT
TAAAAATGTCTGGAACATCTTGGGCAAACGTGTGCGTCAACGTAAACGCCACCAGTCACGCTAGA
TGAACCTGTCCAGGCGTTGGTGGAGAATGGGACAGACTGCATCAATTACCATAAGTAGACTCATT
TGCAGCGAATCAGTCAGTGTGTTGACCAATAACGGGGGCATTACGCACTACTGACGCAAAACAATGT
CAATTTCCGTTTCTTACCCATTCCCTTCTTTCACGGACCATAACAGCAAGAGAACTGNTTAGGTAA
TGAAATACCGGTGAATTATTGTTAACTGGATTCCCTTCTTGTAAAGATAACAATTAGTTTGGGACCA
ATTATTATTATCATTAGTTTGTATTGACCTTGAAATTCGAAGTTCCTCTACATTTTTTAAGGAGT
TTATTTGATTGACAATGAAATGTAAGAAAAGAGCAAATCGTAAAATACGTTAAAAATTATTCCTTA
AACATCAGTCTCTAACTTCAGTTTAAATTGCCAGTAACACGTGTTATATGATGTTTCCGTTTCTCT
TTGTTTTTTTAGCATTCAACTTATTTGATATAACGTTTTACTGTTTTAGATTACATCAAACCTGCAG

DOMAIN 2E

ATGGGCTTTCACAACATAATCTTGTGCGAAAAGAAGTAAGCTCTCTTACAACACTGGAGAAACATT
TTTTGAGGAAAGCTCTCAAGAACATGCAAGCAGATGATTCTCCAGACGGATATCAAGCTATTGCTT
CTTTCCACGCTTTGCCTCCTCTTTGTCCAAGTCCATCTGCTGCACATAGACACGCTTGTTGCCTCC
ATGGTATGGCTACCTTCCCTCAGTGGCACAGACTCTACACAGTTCAGTTCGAAGATTCTTTGAAAC
GACATGGTTCTATTGTGCGGACTTCCATATTGGGATTGGCTGAAACCGCAGTCTGCACTCCCTGATT
TGGTGACACAGGAGACATACGAGCACCTGTTTTACACAAAACCTTCCCAAATCCGTTCTCAAGG
CAAATATAGAATTTGAGGGAGAGGGAGTAACAACAGAGAGGGATGTTGATGCTGAACACCTCTTG
CAAAAGGAAATCTGGTTTACAACAACCTGGTTTTGCAATCAGGCACTATATGCACTAGAACAAAGAAA
ATTACTGTGACTTTGAAATACAGTTCGAAATTTGCATAATGGAATTCATTATGCGGTGGAGGAT
CAAAGACCCATTCAATAGGTCATCTTCATTACGCATCATACGATCCACTGTTCTATATCCACCATT
CGCAGACAGATCGCATTTGGGCTATCTGGCAAGCTCTCCAGGAGCACAGAGGTCTTTCAGGGAAGG
AAGCACACTGCGCCCTGGAGCAAATGAAAGACCCTCTCAAACCTTTCAGCTTTGGAAGTCCCTATA
ATTTGAACAAACGCACTCAAGAGTTCTCCAAGCCTGAAGACACATTTGATTATCACCGATTCCGGT
ATGAGTATGATTCCCTCGAATTTGTTGGCATGTCTGTTTCAAGTTTACATAACTATATAAAACAAC
AACAGGAAGCTGATAGAGTCTTCGCAGGATTCCTTCTTAAAGGATTTGGACAATCAGCATCCGTAT
CGTTTGATATCTGCAGACCAGACCAGAGTTGCCAAGAAGCTGGATACTTCTCAGTTCTCGGTGGAA
GTTTCAGAAATGCCGTGGCAGTTTGACAGGCTTTACAAGTACGACATTACAAAACGTTGAAAGACA
TGAAACTGCGATACGATGACACATTTACCATCAAGGTTACATAAAGGATATAGCTGGAGCTGAGT
TGGACAGCGATCTGATTCCAACCTCCTTCTGTTCTCCTTGAAGAAGGAAAGC

INTRON 2E/2F (SEQ ID NO:150)

GTATGTATCTCATGTTTCTCAAATAATTTGATTTTCAATGCCCTTACTATAAAGCACAGTTATTGT
TCAGTGCCAGTAACCGTTTATTTACGTAAATGTTACAGGCTATTATAATCAAAAATACATTACCGA
TATTGTTTACCACACAATTATATCATTTGTCAAATCTACCCCCATTACCTGCGTTTTGAATTTGTA
ACCTTCTGACAAAAATGAATTAGCAAGAGCTCTGATGAAGAACATAATGAACAACACCTATCTTTC
TTCTTTCAATGACGGTTTAAACAATACAATGCACAATGTAAAAAATATATATATATATATAATTTT
ATATCTACAGTTAATGCAATGACTCCACTAATTCAGGGAAACACATTTTCAG

DOMAIN 2F-1 (1st part of domain f)

ATGGGATCAATGTACGTCACGTTGGTCGTAATCGGATTCGTATGGAACCTATCTGAACTCACCGAGA
GAGATCTCGCCAGCCTGAAATCTGCAATGAGGTCTCTACAAGCTGACGATGGGGTGAACGGTTATC
AAGCCATTGCATCATTCCACGGTCTCCCGGCTTCTTGTGATGATGAGGGACATGAG

INTRON 2F (SEQ ID NO:151)

GTAAATAAAACGTCCAGTCATCGGAAACCCGCCAGATATATGGGTTTTTTTCTATTTAAACAAA
AAAGCAGAGACAAAAAGATTATTTAAAGTCACATTTAACTTGATATCAGATCAATAGTTTGGCTAG
TTAGTGCTCTATATCCCTCAAATCCTTCGAATCTTTAAGCCTCGTGATATTTTGACAAACAGAGAA
GACTTAGTAGCCCAGACTTTCCCTTATTTTTTCTGAAAATCTTAATACGGATATTAAATGGATTC
ATTCTGCAACCTACAACCATAGCCCATATGTTATTATTTTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATTGCCTGTTGTATCCACGGAATGCCAGTATTCCCACACTGGCACAGGCTTTACACCCTGCAAATG
GACATGGCTCTGTTATCTCACGGATCTGCTGTTGCTATTCCATACTGGGACTGGACCAAACCTATC
AGCAAACCTGCCTGATCTCTTCACCAGCCCTGAATATTACGATCCTTGGAGGGATGCAGTTGTCAAT
AATCCATTTGCTAAAGGCTACATTAAATCCGAGGACGCTTACACGGTTAGGGATCCTCAGGACATT
TTGTACCACTTGCAGGACGAAACGGGAACATCTGTTTTGTTAGATCAAACCTCTTTTAGCCTTAGAG
CAGACAGATTTCTGTGATTTTGAGGTTCAATTTGAGGTCGTCCATAATGCTATTCACTACTTGGTG
GGTGGTCGACAAGTTTATGCTCTTTCTTCTCAACACTATGCTTCATATGACCCAGCCTTCTTTATT
CATCACTCCTTTGTTGACAAAATATGGGCAGTCTGGCAAGCTCTGCAAAGAAGAGAAAGCGTCCC
TATCATAAAGCGGATTGTGCTCTTAACATGATGACCAAACCAATGCGACCATTTCACACGATTTTC
AATCACAATGGATTACAAAAATGCACGCAGTCCCCAACACTCTATTTGACTTTCAGGACCTTTTC
TACACGTATGACAACTTAGAAATTGCTGGCATGAATGTTAATCAGTTGGAAGCGGAAATCAACCGG
CGAAAAAGCCAAACAAGAGTCTTTGCCGGGTTCTTCTACATGGCATTGGAAGATCAGCTGATGTA
CGATTTTGGATTTGCAAGACAGCTGACGACTGCCACGCATCTGGCATGATCTTTATCTTAGGAGGT
TCTAAAGAGATGCACTGGGCCTATGACAGGAACTTAAATACGACATCACCCAAGCTTTGAAGGCT
CAGTCCATACACCCTGAAGATGTGTTTGACACTGATGCTCCTTTCTTCATTAAAGTGGAGGTCCAT
GGTGTAACAAGACTGCTCTCCCATCTTCAGCTATCCAGCACCTACTATAATCTACTCAGCTGGT
GAAG

INTRON 2F-2/2G (SEQ ID NO:152)

GTGAGAGAACTATAATAGTGTATGTCGGCAAAAAATGTGCTCATATCATGACTCTGTTGGCCGGT
GGTTGCTCTCCTCTCCTCCTCCACCACCACCGGTACCTCCACCTGTCAGGGCATCAATGTACCATG
AAAATGTCTACAATACTAGGCCTCCTGTAGAAGCACGTAAGATTTACATGGCCGGTTTGTAACCTAG
TTTAAAGTGCTTCACAGTAACCAAACAGTCTCTAAAGATTAATGTCTGTTTAAATTTAATGCC
ACATTTTCAACTGACATATTCTTGCAATTAAGTACAAATGAAGTAGTATAAATTATCCACAAATAG
CGTGATGCACCACAAATATAAACCGAGTGCTTTTTTTGGCATTCCCCACTTGTTCTGGCATGATCAC
ATCATAGATCTCGTTCATGAAGATACTGTTGGATGCTTTTTTCCCAATATGCCCAATCTGTTAAAT
TATTTACACGACCGCAGTGTGTACTTTCACTCAGATCTTTACAATGTGTTTGTAACGTTTACA
ATTAGCGTTATGATTGAAATATTACCCCTGCTACGTTAAATCACATTCACCTCACTCATCTGATGT
ACTTTACAGGTCATACCGATGATCACGGCTCAG

DOMAIN 2G-1 (1st part of domain g)

ATCATATTGCTGGCAGTGGAGTCAGGAAAGACGTGACGTCTCTTACCGCATCTGAGATAGAGAACC
TGAGGCATGCTCTGCAAAGCGTGATGGATGATGATGGACCCAATGGATTCCAGGCAATTGCTGCTT
ATCACGGAAGTCCTCCCATGTGTCACATGCCTGATGGTAGAGACGTTGCATGTTGTACTCATG

INTRON 2G-1/2G-2 (SEQ ID NO:153)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAACAAAAATGTTTTA
TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-2 (2nd part of domain g)

GAATGGCATCTTTCCCTCACTGGCACAGACTGTTTGTGAAACAGATGGAGGATGCACTGGCTGCGC
ATGGAGCTCACATTGGCATAACCATACTGGGATTGGACAAGTGCGTTTAGTCATCTGCCTGCCCTAG
TGA CTGACCACGAGCACAATCCCTTCCACCAC

INTRON 2G-2/2G-3 (SEQ ID NO:154)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAAATGTTT
TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATAACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTT
AG

DOMAIN 2G-3 (3rd part of domain g)

GGACATATTGCTCATCGGAATGTGGATACATCTCGATCTCCGAGAGACATGCTGTTCAATGACCCC
GAACACGGGTCAGAA TCATTCTTCTATAGACAGGTTCTCTTGGCTCTAGAACAGACAGACTTCTGC
CAATTTGAAGTTCAGTTTGAAATAACACACAATGCAATCCACTCTTGGACTGGAGGACATACTCCA
TATGGAATGTCATCACTGGAATATACAGCATATGATCCACTCTTTTATCTCCACCATTCCAACACT
GATCGTATCTGGGCCATCTGGCAGGCACTCCAGAAATACAGAGGTTTTCAATACAACGCAGCTCAT
TGCGATATCCAGGTTCTGAAACAACCTCTTAAACCATTGAGCGAGTCCAGGAATCCAAACCCAGTC
ACCAGAGCCAATTCTAGGGCAGTCGATTTCATTTGATTATGAGAGACTCAATTATCAATATGACACA
CTTACCTTCCACGGACATTCTATCTCAGAACTTGATGCCATGCTTCAAGAGAGAAAGAAGGAAGAG
AGAACATTTGCAGCCTTCCTGTTGACGGATTGGCGCCAGTGCTGATGTTTCGTTTGATGTCTGC
ACACCTGATGGTCATTGTGCCTTTGCTGGAACCTTCGCGGTA CTGGTGGGGAGCTTGAGATGCCC
TGGTCCTTTGAAAGATTGTTCCGTTACGATATCACAAAGGTTCTCAAGCAGATGAATCTTCACTAT
GATTCTGAGTTCACCTTTGAGTTGAAGATTGTTGGCACAGATGGAACAGAACTGCCATCGGATCGT
ATCAAGAGCCCTACCATTGAACACCATGGAGGAG

INTRON 2G/2H (SEQ ID NO:155)

GTATGTTTTGAGATCCACATAATCTTCTACCCTGTCTCATTTCTAATGCTCTTCAATACACAATTT
ATATAGCCTTTGAGCTTCAGATGTATTACGGACAGGCATTACAGTATACATGTAATATGGTTTTCT
GCTATTTGCAAAAATTGTGTCCTATCTCTGTTTCAGATCATCATGGCGGTGACACCTAG

DOMAIN 2H (SEQ ID NO:159)

GTCACGATCACAGTGAACGTCACGATGGATTTTTTCAGGAAGGAAGTCGGTTCCCTGTCCCTGGATG
AAGCCAATGACCTTAAAAATGCACTGTACAAGCTGCAGAATGATCAGGGTCCCAATGGATATGAAT
CAATAGCCGGTTACCATGGCTATCCATTCTGCCCCTGAACATGGTGAAGACCAGTACGCATGCT
GTGTCCACGGAATGCCTGTATTTCCACATTGGCACAGACTTCATACAATCCAGTTTGAGAGAGCTC
TCAAAGAACATGGTTCTCATTTGGGTCTGCCATACTGGGACTGGAC

Figure 11

Primary structure of the KLH2 protein

DOMAIN B

GLPYWDWTMPMSHLPELATSETYLDPVTGETKNNPFHHAQVAFENGVTSRNPDAKLFMKPTYGDHT
YLFDSMIYAFEQEDFCDFEVQYELTHNAIHAWVGGSEKYSMSSLHYTAFDPIFYLLHHSNVDRWLAI
WQALQIRRGKSYKAHCASSQEREPLKPFASFSSPLNNNEKTYHNSVPTNVYDYVGVLLHYRYDDLQFG
GMTMSELEEYIHKQTQHDRTFAGFFLSYIGTSASVDIFINREGHDKYKVGSFVVLGGSKEMKWGFD
RMYKYEITEALKTLNVAVDDGFSITVEITDVGSPPSADLIPPAIIFDVVR

DOMAIN C

ADAKDFGHSRKIRKAVDSLTVEEQTSLLRRAMADLQDDKTSGGFQQIAAFHGEPKWCPSPAEKKA
CCVHGMVFPWHRLTYVQGENALRKHGFTGGLPYWDWTRPMSALPHFVADPTYNDSVSSLEEDNP
WYHGHDIVSGHDTTRAVRDDLYQSPGFQHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGG
NEPYSMSSLRYTTYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGLDSVI
NPDDETREHSVPFRVFDYKNNFDYYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIGQSAL
VKFYVCKHNVSDCDHYAGEFYILGDEAMPWRYDRVYKYEITQQLHDLDLHVGDNFFLKYEAFDLN
GGSLGGSIFSQPSVIFEPAA

DOMAIN D

GSHQADEYREAVTSASHIRKNIRDLSERGEIESIRSAFLQIQKEGIYENIAKFHGKPGLCHEHDGHPV
ACCVHGMPTFPWHRLTYVLQVENALLERGSAAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPN
PFFRGHIAFENAVTSRDPQPELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYS
LSSLDYTAFDPVFFLHHANVDRIWAIWQDLQRYRKKPYNEADCAVNEMRKPLQPFNNPELNSDSMT
LKHNL PQDSFDYQNRFRYQYDNLQFNHFSIQKLDQTIQARKQHDRVFAFGFILHNIGTSASVVDIYIC
VEQGGEQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDGHGFDIKVDVRAVNGSH
LDQHILNEPSLLFVPGERKNIYY

DOMAIN E

DGLSQHNLVRKEVSSLTTLLEKHFLRKALKNMQADDS PDGYQAIASFHALPPLCPSPSAHRHACCL
HGMATFPQWHRLTYVQFEDSLKRHGSIVGLPYWDWLKPQSALPDLVTQETYEHLFSHKTFPNPFLK
ANIEFEGEGVTTERDVAEHLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWVGG
SKTHSIGHLHYASYDPLFYIHHSQTDRIWAIWQALQEHRLSGKEAHCALEQMKDPLKPFSGSPY
NLNKRTOEFSKPEDTFDYHRFGYEYDSLEFVGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV
SFDICRPDQSCQEAGYFVSVLGGSSSEMPWQFDRLYKYDITKTLKDMKRLYDDTFTIKVHIKDIAGAE
LSDSLIPTPSVLLEEGK

DOMAIN F

HGINVRHVGRNRIRMELSELTERDLASLKSAMRSLQADDGVNGYQAIASFHGLPASCHDDEGHEIA
CCIHGMPVFPWHRLTYLQMDMALLSHGSAVAIPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNP
FAKGYIKSEDAYTVRDPQDILYHLQDETGTSLVLLDQTLLEQTDFCDFEVQFEVVHNAIHYLVGG
RQVYALSSQHYASYDPAFFIHHSFVDKIWAVWQALQKKRKPYPHKADCALNMMTKPMRPFADFNH
NGFTKMHAVPNTLFDQDLFYTYDNLEIAGMNVNQLAEINRRKSQTRVFAGFLLHGIGRSADVRF
WICKTADDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFDTDAPFFIKVEVHGV
NKTALPSSAIPAPTIISAGE

DOMAIN G

DHIAGSGVRKDVTSLTASEIENLRHALQSVMDDDGPNGFQAIAYHGSPPMCHMPDGRDVACCTHG
MASFPHWHRLFVKQMEDALAAHGAHIGIPYWDWTSASFHLPALVTDHEHNPFFHHGHIAHRNVDTSR
SPRDMLFNDPEHGSESFFYRQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYD
PLFYLLHNSNTDRIWAIWQALQKYRGFYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFD

YERLNYQYDTLTFHGHSELDAMLQERKKEERTFAAFLLHGFASADVSDVCTPDGHCAFAGTF
AVLGGELEMPWSFERLFRYDITKVLKQMNLHYDSEFHFELKIVGTDGTELPDRIKSPTIEHHGG

DOMAIN H (SEQ ID NO:158)

GHDHSEKHDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYESIAGYHGYPFLCPEHGEDQYAC
CVHGMPVFPHWRLHTIQFERALKEHGSHLGLPYWDW